CYBERINFRASTRUCTURE SUPPORTING WATERSHED HEALTH MONITORING AND MANAGEMENT

TONY B. SZWILSKI, JACK SMITH, JUSTIN CHAPMAN & MARK LEWIS
Centre for Environmental, Geotechnical and Applied Sciences, Marshall University, USA

ABSTRACT
A distributed cyberinfrastructure (CI) is being built to collect, manage and analyse watershed health data in the Appalachian region (USA), with an early emphasis on the Fourpole Creek watershed near Marshall University, the subject of a TMDL (Total Maximum Daily Load) study by the US EPA in 2002. The CI architecture comprises an interactive web-based portal (*Aquavit*) supported by collaboration tools, modelling tools, high-performance computing (HPC), a federated data management system providing batch import/export capabilities for field-collected lab-analysed sample data or UAV-collected post-processed data, and continuous ingestion of real-time data from a network of sensors. Spatial and temporal data collected and analysed locally includes rain intensity, temperature, solar radiation, soil moisture, wind velocity, sediment properties, and several water quality parameters. Additional watershed health indicators are collected or derived from external sources, such as hydrography and hydrology data, land use data, soil types, land cover types, point and non-point contaminant sources, weather data, climate model data, multi-spectral satellite imagery.
Keywords: cyberinfrastructure, watershed health, water quality, Appalachia, data management, sensor networks, UAVs, land use, climate change, modelling.

1 INTRODUCTION
The Appalachian watersheds in the United States are a source of significant freshwater, and the need for securing these resources to support future economic development is crucial as the demand for freshwater continues to grow. Thus, it is essential to maintain a healthy watershed that can respond to environmental stresses, such as changes in land use and weather patterns due to climate change and be capable of filtering pollutants sufficiently enough to surpass all water quality standards. However, this region faces unique challenges that includes the mineral extraction industry and untreated wastewater effluent discharges.

   This study and its earliest cyberinfrastructure (CI) development has focused initially on the Fourpole Creek watershed (HUC 050901011006), near Huntington, WV (see Fig. 1). This watershed was the subject of a TMDL study by the US EPA in 2002 for Aluminium and Faecal Coliform [1]. Fourpole Creek is also a priority area of a Long-Term Control Plan (LTCP) of Combined Sewage Overflows (CSOs) developed by the Huntington Sanitary Board (HSB) in 2010 [2]. Additionally, a community organization, the Fourpole Creek Watershed Association (FCWA) [3], exists to help oversee and promote the health of this watershed and has ties to Marshall University, as well as to most of the agencies with more formal oversight in the region.

1.1 Need for improved understanding at the watershed level

We need to improve our ability to detect contaminants and better understand their transport characteristics, complex chemical interactions, impact on biological processes, chemical–physical interaction with sediments, and their ultimate fate at the watershed and sub-watershed level [4]. We also need to better account for the impacts of land use, storm water management, and climate change on sediment and contaminant transport.
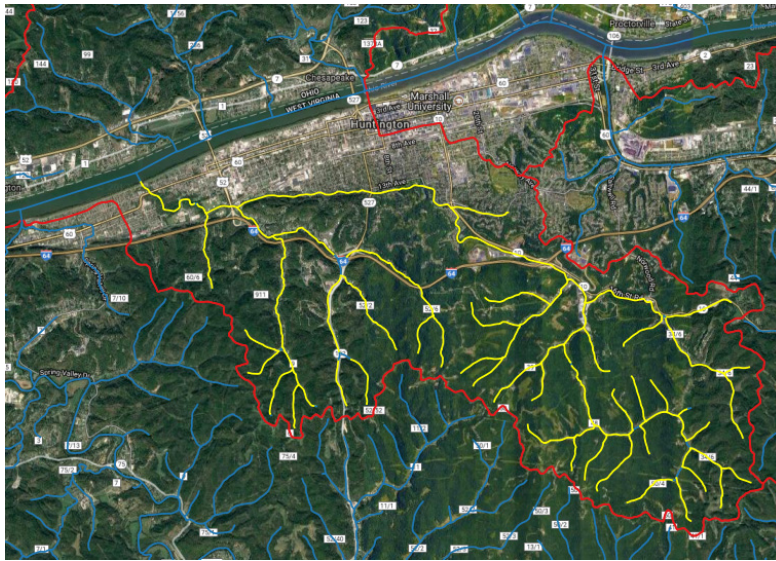
Figure 1: Fourpole Creek watershed.

1.1.1 Impact of land cover and use

One primary factor affecting the overland flow of surface water and channelling it to other water bodies (streams, rivers, lakes, reservoirs, etc.) or infiltrating it into soil and underground water systems (aquifers) is land cover and use which can vary significantly across a studied area. Zoning and regulations regarding agricultural and industrial practices help mitigate some effects, but detailed computer modelling and simulation is usually the only way to develop and test best practices. Soil type also plays an important role in determining overland flow characteristics and best land use practices, especially with regard to sediment transport (erosion) and the types of vegetation and development that can be supported.

1.1.2 Impact of storm water management

In developed areas, runoff is typically redirected to man-made drainage systems that utilize man-made catchments (ponds, reservoirs) and/or natural water systems. Extreme weather events can overwhelm these systems, especially when poorly maintained or inadequately upgraded as development grows. In such areas, land cover or soil type is generally not suitable to handle flow, leading to severe erosion and/or flooding. Furthermore, when these storm sewer systems are combined with wastewater sewer systems, a Combined Sewer System (CSS), overflow (CSO) events can occur leading to untreated waste water bypassing water treatment facilities and entering natural streams and lakes. These extreme events can disrupt controlled discharge of point source contaminants and expectant runoff of non-point source contaminants, which then mixes with the untreated water.

1.1.3 Impact of climate change

Global rainfall distribution is changing, whereby some regions are experiencing more rain and others less [5]. Various studies indicate that precipitation event characteristics are influenced by global climate change, including extreme weather conditions that are most

likely affecting watershed health and water quality, such as increased frequency and intensity of CSO events [6]. Changes in temperature, humidity, and accessible solar radiation (due to atmospheric photochemical reactions, cloud cover, foliage) can all impact the biological and chemical processes that contribute to the fate of contaminants. Changes in wind characteristics can impact erosion, airborne contaminant distribution, oxygen dissolution in water systems, and the whole evapotranspiration cycle.

1.1.4 Impact on sediment and contaminant transport

Higher frequency and intensity of precipitation events result in greater soil sediment dislocation, channel erosion, and increased levels of sediment and contaminants being transported. Any changes in precipitation intensity, rate of surface water runoff, infiltration (soil saturation), and stream flow directly impact the transportation and fate of contaminants and water quality [7]. The climate affects both the temporal and spatial distribution of precipitation (and other weather parameters), and as a consequence, the temporal and spatial redistribution of sediment and associated contaminants. Sediment is the largest undesired constituent of surface water by weight and volume. The direct impact of sediment on aquatic environments has been an issue for many years. Earlier research found sediment intrusion into streambeds severely impacted freshwater mussels living in gravel-bedded and sand-bedded channels [8]. Sediments may include silt, clay, sand, gravel, and organic matter. The fate of sediment-bound contaminants and nutrients requires an understanding of the transportation and deposition of sediment [9]. The interaction between contaminants and sediments in a river can be complex, the physical and chemical properties of each dictates their mass behaviour. Generally, the rate of sediment discharge directly correlates to the level of water discharge across a section of the stream or river [11]. Physical process-based watershed models, such as SWAT [12], HSPF [13], MIKE SHE [14], SIMWE [16], and others, simulate a wide range of processes: contaminant and water runoff, sediment and contaminant transportation [17].

Watershed geomorphological changes, in the form of river and stream channel erosion (hydromorphological), are a principal cause of sediment discharge in surface waters. Precipitation and temperature characteristics impact river water discharge and, consequently, sediment discharge rate. During low flow conditions the concentration of contaminants is likely to increase and further impact the health of the watershed. Model simulations indicate a pattern of decreasing streamflow in the western US and increasing in the eastern US by the mid-21st century due to climate change [18]. Thus, the east will likely experience increases in suspended solids and contaminants such as phosphorous and nitrogen. Agriculture is the principal source of these nutrients, which can cause problems such as toxic algal blooms and degradation of the aquatic ecosystem [19]. Of the two principal climate change factors impacting nutrient transportation, precipitation intensity is a greater influence than temperature [20]. Initiatives to address the issue of nutrients (phosphorous, nitrogen) washing into the rivers typically include precise surface irrigation and intelligent application of fertilizers on farmlands. Part of optimizing fertilizer use and irrigation is the application of moisture sensors and fixed-wing drones utilizing multi-spectral imaging.

1.2 Need for a national cyberinfrastructure

A national CI for an integrated hydrologic model has been proposed by Duffy at Penn State [21], but its scope is well beyond water quality, and it has not yet been deployed beyond the Susquehanna Shale Hills Critical Zone Observatory [22] or made publicly available as a generic CI adaptable to high-fidelity water quality monitoring and modelling.

One attempt at creating a generic cyberinfrastructure for water research is WaterHUB [23], hosted on MyGeoHub [24], a HUBzero-based [25] multi-tenant portal at Purdue University. However, its focus has been primarily on deploying modelling tools and sharing models within specific disciplines at a project level, with little integration between projects and no sustainable path toward self-hosting or adaptation to regional needs.

Several research initiatives are currently underway, such as the one by the Department of Interior (Bureau of Reclamation) to develop real-time water quality programs, taking a watershed approach to manage water quality [26], but they generally lack the scope or generality to form the basis of a national CI.

Collecting accurate, timely and comprehensive data is a key requirement for continually assessing and managing the watershed health. However, creating a CI that supports all the other aspects of effective decision making, such as remote collaboration, data collection (real-time, batch, and external sources), distributed data management (sharing, publishing, discovering, and federation with other systems), and modelling (with adequate compute, data transfer, and visualization resources), remains a significant challenge [27].

2 CYBERINFRASTRUCTURE

The cyberinfrastructure (CI) leverages existing external systems where possible to exchange data, share compute resources, or interoperate with other services. At the core, the CI provides a collaboration portal, *Aquavit*, which is built on the HUBzero platform from Purdue [25] and acts as a single point of entry. A unique part of this platform is its ability to support tool development and web-based deployment within the portal for private or public use. External to the core are high performance computing (HPC), data, and visualization resources, transparently interfaced to the user. The architecture is shown in Fig. 2.

2.1 *Aquavit*

Aquavit ("water is life") is the name used for both the overall CI and the HUBzero-based core, since all user and programmatic access is through this core. HUBzero is an open source software platform for building collaborative Science Gateways that offer a variety of traditional content management system (CMS) functions. The HUBzero core is being extended to provide a data portal to stream real-time or import batch data and push to a national repository for shared access.

Collaborating researchers or anyone granted permission have access to content based on their group and/or project membership within the CMS, which may include real-time and stored data collected across specific watersheds. Accounts on *Aquavit* are free and easy to create using a self-registration process with a robust role-based project and group permission system allowing for tight control over contributed content.

2.2 Collaboration portal

HUBzero (and thus *Aquavit*) at its core is a Content Management System (CMS) based originally on Joomla! [28], but was extended significantly to meet the needs of NanoHUB [29] to support a computationally intensive community of researchers in nano-scale material science, which was later made available as a generic Science Gateway development platform.

2.3 Modelling portal

The HUBzero platform provides users the ability to develop or upload and publish tools for all collaborating researches and stakeholders. Linux-based tools are developed and deployed
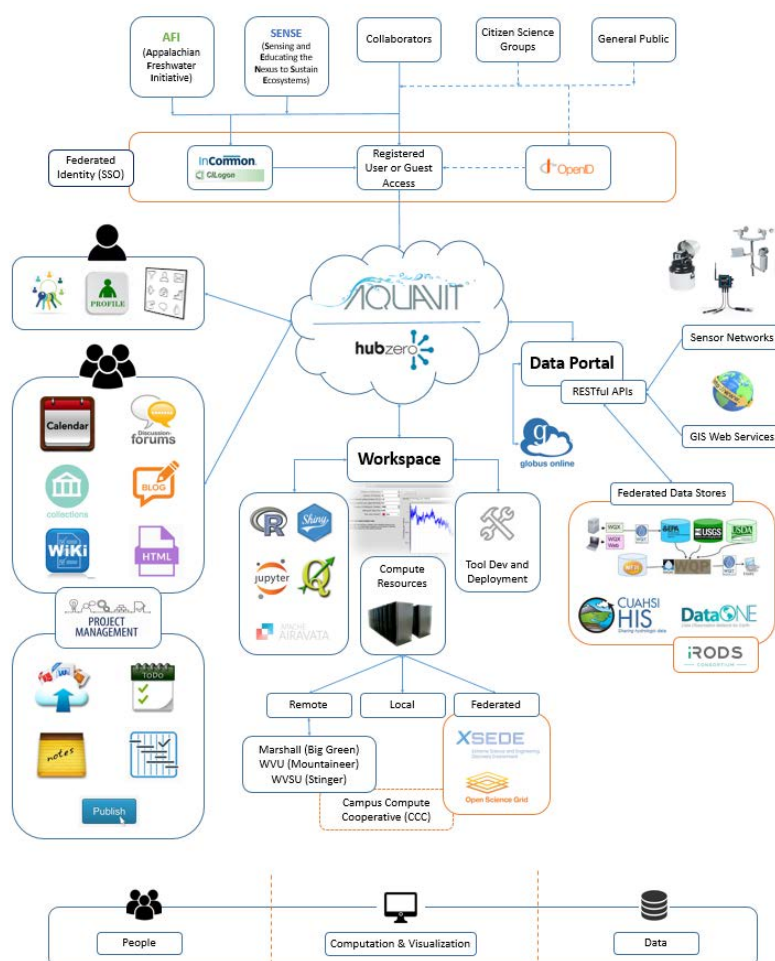
Figure 2: *Aquavit*: HUBzero-based collaboration portal for water research.

within personal virtual containers based on OpenVZ [30]. HUBzero provides the *Rappture* (Rapid application infrastructure) Toolkit [31] to help build web-based GUIs for applications and workflows written in a variety of programming/scripting languages.

Tools deployed on *Aquavit* can run locally in real-time on the *Aquavit* server, an external execution server (coming soon), or in batch mode on the HPC cluster at Marshall University ("Big Green") [32] or other campus clusters (once configured). XSEDE [33] resources, OSG [34] and other federated compute resources are planned additions in the near future [35].

2.4 Data portal

Perhaps the most ambitious enhancement to the HUBzero framework for *Aquavit* is development of a data portal for ingesting batch and real-time water quality data from the field and routinely pushing it to the Water Quality Portal (operated and maintained by the

National Water Quality Monitoring Council for the USGS, US EPA, and USDA) [36], as well as providing local analytics and visualization directly through *Aquavit*. Batch input will be performed by uploading a spreadsheet into a WQX-compatible database on *Aquavit*, after passing it through an import configuration script (XSL stylesheet). The latter can be generated by the Water Quality Exchange (WQX) Web [37] application accessible through the Central Data Exchange (CDX) portal [38] maintained by the STORET [39] team at the US EPA, which requires establishment of CDX account IDs and WQX Organization IDs, which has been done. A WQX-compatible web form (HUBzero Web Component) is also available for manual entry.

Real-time data can be streamed to a REST API that front-ends the WQX-compatible database, essentially an API for the HUBzero Web Component used for manual entry.

2.4.1 External data resources

In addition to a portal for streaming or uploading locally collected data, *Aquavit* will also provide the ability to exchange data with external sources. Some data will be pushed to the WQP, as described above, for broader sharing and pulled down as needed for local modelling efforts. Metadata describing datasets and how to access them will be pushed (published) to DataONE [40] for broader dissemination and discoverability. Datasets, models, and tools (web apps) can be published on HydroShare [41] at CUAHSI [42]. SWAT models, in particular, will be pushed to HydroShare via SWATShare [43], being implemented in *Aquavit* from the WaterHUB project [23].

At the file level, transfer capabilities will be provided by Globus Connect [44] to and from any other Globus endpoint, including personal computers. For distributed file sharing, iRODS [45] is being implemented via iData [46] from the GABBs project [47]. Globus can also facilitate sharing, publication, and discoverability at the file level by also providing appropriate metadata [48].

Several data resources will be made available for download and incorporation into models, such as elevation (3DEP) maps, hydrographic (NHD) datasets, HUC boundaries (WBD), and land cover (NLCD) datasets from The National Map (TNM) provided by the USGS, or MODIS datasets derived from multi-spectral satellite imagery from NASA [50]. Geospatial tools, such as the HydroGlobe Tool [51], will be implemented to help incorporate such data into model building.

3 DATA COLLECTION

Another major component of this effort is collecting field data which will reside in *Aquavit* and/or be posted, through metadata to external sites, for broader sharing and assimilation into larger collections. Data is collected in either real-time, usually from continuously monitoring sensors over the Internet, or in batch fashion, usually as a file uploaded into *Aquavit*. Some data is entered manually through a web form while other data is obtained directly from sensors in the field, with or without any prior QA/QC or intermediate data logging prior to streaming to an *Aquavit* API. Other data comes from lab analysis performed on digital or analog field sample reports: likely manually entered into *Aquavit* by a user. Some data may be complex, such as images or video that have to be post-processed to extract parameters or indices before being entered into *Aquavit*.

3.1 Field stations

Traditional data collection involves visiting a site, collecting grab samples, performing analytical tests, recording results in a notebook, applying QA/QC and submitting a report

with statistical summaries to some researcher or repository. Usually, data from continuously monitored sensors recording to a data logger, downloaded when periodic grab samples are taken, and/or the sensors are serviced. Servicing usually includes recalibration of instruments/sensors and repairs – a manual and time-intensive process. Input to *Aquavit* would involve both analytical sample data results and time-series data collection from data loggers. For most commonly collected water quality parameters, the trend is moving toward more automated sampling and real-time streaming of sensor data, leading to more timely, consistent data, requiring less human intervention and time. However, many ad hoc or infrequent measurements are not routine enough to automate.

3.2 Automated sampling stations

Real-time sensors are gradually taking the place of some laboratory analyses as innovative self-cleaning, self-calibrating, low-energy, digital-readout probes become available, but the need for grab samples and lab analysis will continue for some time. Automated sampling is the next best thing for improving quality, consistency and timeliness, while reducing the amount human intervention and time involved in taking samples. Samplers can now even be remotely controlled to change sample scheduling, respond to events (triggered by on-board sensors), check operation status, etc., further decreasing the need for intervention. Such sampling stations exist in Fourpole Creek. A Teledyne ISCO Automated Sampling Station is automatically collecting grab samples at a fixed stream location collecting depth measurements to estimate flowrate. A home-made sediment trap is deployed nearby. The objective is to gain a better understanding of in-stream sediment, carbon and nutrient processes from low-gradient agricultural watersheds to steep-gradient Appalachian watersheds [53]. This requires periodic manual collection of samples, servicing stations, and lab analysis, and the data uploaded to *Aquavit*.

3.3 Real-time continuous monitoring stations

Ideally, no grab samples need to be taken, lab analysis performed, or regular servicing of equipment required, but this is only possible using sensors that can communicate measurements in real-time. In most field locations, there is no access to power or wired telecommunications, requiring batteries and wireless communications. Eventually the batteries need replaced or certain components replaced/serviced, but sensors can now transmit operational status indicators as well as water quality data, requiring less human intervention. Solar-powered rechargeable batteries and low-power components can dramatically reduce need for routine servicing. Excluding satellite, isolated stations typically require a nearby communication relay station – a frequent challenge with Appalachian topography and infrastructure.

A YSI (EXO2 Multiparameter) Sonde [54] is a good example of a rather self-contained real-time continuous monitoring station, often solar-charged, requiring infrequent servicing, while transmitting data to some receiving station at reasonable frequencies. Such sondes are being deployed by the KY-WV SENSE project (NSF 1632888) [55] in lakes and rivers across Kentucky and West Virginia and, in particular, at strategic locations along the Greenup Pool section of the Ohio River where the Fourpole Creek watershed (and neighbouring watersheds) ultimately drain. This is being done in partnership with ORSANCO [56]. This data will eventually be streamed to *Aquavit* in near real-time using a REST API. A Wireless Davis (Vantage Pro2) Weather Station [57], discussed in the next section on WSNs, when isolated, is another good example.

3.4 Wireless sensor networks

There is little or no access to the Internet in the field, so *ad hoc* networks are needed to wirelessly relay information to an internet gateway. These are usually called Wireless Sensor Networks (WSNs). An affective CI and well-planned WSNs can capture spatial and temporal data in real-time (or near real-time) across a wide area of a watershed [58].

A significant component of this effort is development of energy-efficient sustainable WSNs [60]. The first phase is focussing on creating a mesonet of weather stations and other sensors to study characteristics and impacts of precipitation events on watersheds. Initial focus will be on characterizing extreme, localized, short-duration rain events that provide conditions to produce flash floods across Appalachia.

Solar-powered wireless Davis (Vantage Pro2) weather stations are being deployed on the Marshall campus, at high-schools in the region, and strategic sites within the Fourpole Creek watershed. Currently, stations are reporting all attached sensor data in three-minute time intervals to *Aquavit*: available to researchers on demand. Stations leverage commercial Meteobridge [61] devices to aid data processing and relay to *Aquavit* in partnership with the WV State Climatologist [62], WV National Weather Service (NWS) Forecast Office [63], USGS WV Water Science Center [64], and WV Division of Homeland Security and Emergency Management (DHSEM) [65], who maintains a federally supported network of rain gages across the Appalachian region as part of the Integrated Flood Observing and Warning System (IFLOWS) [66]. IFLOWS was created to support the National Flash Flood Program Development Plan in 1978 [67]. By combining forces, data from 114 weather stations is transmitted and assimilated in near-real time, which includes: rain fall, temperature, humidity, barometric pressure, dew point and ultraviolet light. This is truly a distributed, collaborative, multidisciplinary effort requiring a flexible and open cyberinfrastructure to tie it all together in a federated fashion.

3.5 UAVs

Monitoring the spatial and temporal variation in physical river parameters is important for understanding and improving habitat quality and distribution, especially with respect to potential impacts of climate change [68]. A fixed-wing UAV will be part of a system to assess changes in river and stream hydromorphology (sinuosity) resulting from riverbank erosion. A multispectral sensor attached to the fixed wing UAV will monitor the watershed landscape and habitat. Used primarily for agriculture, the sensor produces sharp and accurate images, is small, and light weight. With better precision than satellite imagery, the sensor can capture images that are sensitive to the biological condition of plants.

4 DISCUSSION

Effective management of a watershed requires the coordination between many stakeholders: university researchers across multiple institutions, local/state/federal agencies, non-profit organizations, watershed groups, citizen scientists, farmers, and anyone with a personal or professional stake in monitoring or managing water quality. Thus, developing CI to collect, share, model and analyse watershed data, and make available to all stakeholders, is a key element of this project. The goal is to facilitate collaboration, data collection, modelling, data and model sharing, and federation with other systems to form a distributed and scalable CI with as few barriers to collaboration and local adaptation as possible.

Modelling a complex watershed, such as Fourpole Creek, takes into account land use options, land coverage, topology, contaminant sources, and water management systems. The

objective is to establish best management practices at the local level and affectively mitigate against expected and unexpected threats to water quality requires data with much higher fidelity, both spatially and temporally, than is usually available for a watershed. A good part of this effort is to gather and manage much more data from a variety of sources, in different manners, at different time and spatial scales, for a diverse set of parameters – as efficiently as possible. To do this intelligently requires effective use of different technologies. Managing those technologies in a holistic way requires an underlying CI, which is at the heart of this project.

An upcoming feature of HUBzero (currently in beta testing) is the ability to deploy web applications ("weber apps") outside the OpenVZ container (through a proxy service) on the host or external web server. This will be particularly useful for deploying R Shiny [69] and Jupyter (Python) Notebook [70] applications.

5 CONCLUSIONS

Building CI infrastructure to monitor the health of watersheds is an important and long-term endeavour. Adapting new technologies to facilitate this task is a slow process that requires a continuous investment and efficient utilization of resources, and keen and effective collaboration from the principal stakeholders. The principal challenge is to develop a practical system that adds value, is sustainable, replicable (modular, extensible, open source), and directed toward true federation in terms of standards and APIs.

REFERENCES

[1]   US EPA, Region 3. Total Aluminum and Fecal Coliform Bacteria TMDLs for the Fourpole Creek, West Virginia, 2002. dep.wv.gov/WWE/watershed/TMDL/grpe /Documents/Lower%20Ohio/2977_fourpole_tmdl.pdf.

[2]   Huntington Sanitary Board. CSO Long Term Control Plan, 2010. www.huntingtonsb.com/pages/pdfs/NPDESApprovedLTCP.pdf.

[3]   Fourpole Creek Watershed Association. fourpolewatershed.wordpress.com.

[4]   Karr, J.R., Defining and Measuring River Health. *Freshwater Biology*, **41**(2), 1999.

[5]   NASA, www.nasa.gov.

[6]   US EPA, National Pollutant Discharge Elimination System (NPDES): Combined Sewer Overflows (CSOs). www.epa.gov/npdes/combined-sewer-overflows-csos.

[7]   Whitehead, P.G., Wilby, R.L., Battarbee, R.W., Kernan, M. & Wade, A.J., A Review of the Potential Impacts of Climate Change on Surface Water Quality. *Hydrological Sciences Journal*, **54**(1), pp. 101–123, 2009.

[8]   Castro, J. & Reckendorf, F., Effects of Sediment on the Aquatic Environment: Potential NSCS Actions to Improve Aquatic Habitat – Working Paper No. 6, 1995.

[9]   Koltun, G., Landers, M.N., Nolan, K.M. & Parker, R.S., Sediment transport and geomorphology issues in the water resources division. *Proceedings of the U.S. Geological Survey (USGS) Sediment Workshop*, 1997.

[10]   US EPA, Contaminated Site Cleanup Information (CLU-IN). clu-in.org/ contaminantfocus/default.focus/sec/Sediments/cat/Overview/.

[11]   Colby, B., Relationship of sediment discharge of stream flow, 1959. pubs.usgs.gov/of/ 1956/0027/report.pdf.

[12] Arnold, J., Williams, A., Srinivasan, R., King, B. & Griggs, A., Soil and Water Assessment Tool (SWAT). swat.tamu.edu.

[13] USGS. Hydrological Simulation Program – Fortran (HSPF). water.usgs.gov/software/HSPF/.

[14] DHI, DHI MIKE SHE. www.mikepoweredbydhi.com/products/mike-she.

[15] Mitas, L. & Mitasova, H., Distributed soil erosion simulation for effective erosion prevention. *Water Resources Research*, **34**(3), pp. 505–516, 1998.

[16] GRASS 7, r.sim.water (SIMWE). grass.osgeo.org/grass74/manuals/r.sim.water.html.

[17] Mitasova, H., Mitas, L. & Brown, W., Multiscale simulation of land use impact on soil erosion and deposition and deposition patterns. *Proceedings of the 10th International Soil Conservation Organization Meeting*, USDA-ARS National Soil Erosion Research Laboratory, Purdue University, 1999.

[18] Johnson, T. et al., Modeling Streamflow and Water Quality Sensitivity to Climate Change and Urban Development in 20 U.S. Watersheds. *Journal of American Water Resources Association (JAWRA)*, **51**(5), pp. 1321–1341, 2015.

[19] Carpenter, S.R., Caraco, N.F., Correll, D.L., Howarth, R.W., Sharpley, A.N. & Smith, V.H., Nonpoint Pollution of Surface Waters with Phosphorous and Nitrogen. *Ecological Applications*, **8**(3), pp. 559–568, 1998.

[20] McIntosh, B., With Droughts and Downpours, Climate Change Feeds Chesapeake Bay Algal Blooms, 2016.

[21] Duffy, C., Leonard, L., Bhatt, G., Yu, X. & Giles, L., Watershed reanalysis: towards national cyberinfrastructure for model data integration (NSF EAR 0725019). In workshop on measuring the impact of e-science. *7th IEEE International e-Science Conference*, Stockholm, Sweden, 2011.

[22] Susquehanna Shale Hills Critical Zone Observatory (CZO), criticalzone.org/shale-hills/.

[23] WaterHUB. mygeohub.org/groups/water-hub.

[24] MyGeoHub, mygeohub.org.

[25] Purdue University, HUBzero. hubzero.org.

[26] Bureau of Reclamation, Department of Interior, Cyber-Infrastructure Development in Support of Real-Time Water Quality Management (BOR-MP-17-N001), 2017.

[27] Muste, M.V. et al., End-to-End Cyberinfrastructure for Decision-Making Support in Watershed Management. *Journal of Water Resources Planning and Management*, **139**(5), pp. 565–574, 2013.

[28] Joomla! www.joomla.org.

[29] NanoHUB, nanohub.org.

[30] OpenVZ, openvz.org/Main_Page.

[31] Rappture Toolkit, nanohub.org/infrastructure/rappture/.

[32] Marshall University, Big Green (HPC Cluster). www.marshall.edu/itresearch/resources-and-tools/biggreen-cluster/.

[33] XSEDE, www.xsede.org.

[34] Open Science Grid (OSG), www.opensciencegrid.org.

[35] Govindaraju, R.S. et al., Vision of Cyberinfrastructure for End-to End Environmental Exploration (C4E4). *Journal of Hydraulic Engineering*, **14**(1), 2009.

[36] Water Quality Portal (WQP), www.waterqualitydata.us.

[37] EPA Water Quality Exchange (WQX), www.epa.gov/waterdata/water-quality-data-wqx.

[38] EPA Central Data Exchange (CDX), cdx.epa.gov.

[39] EPA STOrage and RETrieval (STORET) Data Warehouse, www.epa.gov/waterdata/concepts-and-definitions-storage-and-retrieval-and-water-quality-exchange.

[40] Data Observatory Network for Earth (DataONE), www.dataone.org.

[41] CUAHSI, HydroShare. www.hydroshare.org.

[42] Consortium of Universities for the Advancement of Hydrologic Science, Inc (CUAHSI), www.cuahsi.org.

[43] SWATShare, mygeohub.org/groups/water-hub.

[44] Globus Connect, www.globus.org/globus-connect.

[45] iRODS, irods.org.

[46] GABBs, iData. mygeohub.org/groups/gabbs/aboutidata.

[47] GABBs, mygeohub.org/groups/gabbs.

[48] Globus Transfer, Share, Publish and Build, www.globus.org.

[49] USGS, The National Map (TNM). nationalmap.gov.

[50] NASA, Moderate Resolution Imaging Spectroradiometry (MODIS). modis.gsfc.nasa.gov.

[51] Rajib, A., Merwade, V., Zhao, L., Shin, J., Smith, J. & Song, C., HydroGlobe Tool, 2017. mygeohub.org/resources/hydroglobetool.

[52] International Atomic Energy Agency (IAEA), www.iaea.org.

[53] Ford, W., Personal communication, 2016–2017.

[54] YSI, EXO2 Sonde. www.ysi.com/EXO2.

[55] KY-WV SENSE Project (NSF 1632888), www.nsf.gov/awardsearch/showAward?AWD_ID=1632888.

[56] Ohio River Valley Water Sanitation Commission (ORSANCO). www.orsanco.org.

[57] Davis, Wireless Vantage Pro2 Weather Station. www.davis.com/Assets/MoreInfo/86403-13Specs.pdf.

[58] Faustine, A., Mvuma, A.N., Mongi, H.J., Gabriel, M.C., Tenge, A.J. & Kucel, S.B., Wireless sensor networks for water quality monitoring and control within Lake Victoria Basin: prototype development. *Wireless Sensor Network*, **6**(12), pp. 281–290, 2014.

[59] VanBriesen, J., RAPID: Cyber-infrastructure Development for Dunkard Creek Watershed (NSF #1001400), 2008. www.nsf.gov/awardsearch/showAward?AWD_ID=1001400.

[60] Malik, H. & Szwilski, A., Towards monitoring the water quality using hierarchal routing protocol for wireless sensor networks. *Proceedings of the 7th International Conference on Emerging Ubiquitous Systems and Pervasive Networks (EUSPN)*, 2016.

[61] Meteobridge, wiki.meteobridge.com/wiki/index.php/Home.

[62] WV State Climate Office and Meteorology, www.marshall.edu/met/.

[63] National Weather Service (NWS) Forecast Office, Charleston, WV, w2.weather.gov/climate/index.php?wfo=rlx.

[64] USGS WV Water Science Center, wv.usgs.gov/about/employeedirectory.html.

[65] WV Division of Homeland Security and Emergency Management (DHSEM).

[66] Integrated Flood Observing and Warning Systems (IFLOWS). www.vaemergency.gov/emergency-management-community/emergency-management-resources/integrated-flood-observing-and-warning-systems-iflows/.

[67] US Department of Commerce, National Flash Flood Program Development Plan FY 1979–1984, 1978.

[68] Woodget, A.S., Austrums, R., Maddock, I.P. & Habit, E., Drones and digital photography: from classifications to continuums for monitoring river habitat and hydromorphology. *WIREs: Water*, Jul./Aug., **4**(4), p. 1222, 2017.

[69] R Studio, R Shiny. shiny.rstudio.com.

[70] Jupyter Notebook, jupyter.org.