Short communication

The Mid-Atlantic Watershed Atlas (MAWA): Open access data search & watershed-based community building

Patrick M. Reed a,⁎, Brian Bills b, Michael Anderson b, Blake Ketchum b, Michael Piasecki c

a Department of Civil and Environmental Engineering, The Pennsylvania State University, 212 Sackett Building, University Park, PA 16802, USA
b Center for Environmental Informatics, The Pennsylvania State University, 2217 Earth-Engineering Sciences Building, University Park, PA 16802, USA
c Department of Civil & Architectural Engineering, Drexel University, 3141 Chestnut Street, Philadelphia, PA 19104, USA

Article info

Article history:
Received 30 April 2009
Received in revised form 10 December 2009
Accepted 13 December 2009
Available online xxx

Keywords:
Informatics
Information services
Watershed management
Visualization
Geographic information system

Data/Software availability

http://www.mawa.psu.edu.

1. Introduction

The Mid-Atlantic Watershed Atlas (MAWA) initiative represents a long-term effort to establish a regionally focused virtual organization that facilitates data access, data exploration, and broad collaboration across the northeastern United States (see Fig. 1). As noted by Reed et al. (2006), the past 400 years of human development has strongly influenced the evolution of the quantity and quality of the Mid-Atlantic's regional hydrologic and ecological services. The Mid-Atlantic's topographical, land-cover, and geological heterogeneity has strongly influenced coupling and feedbacks between the human and natural systems that have evolved to define the region's watershed systems (Goetz et al., 2004). Challenges posed to the region's watersheds are strongly representative of the growing global concerns over the emergent dependencies between population pressure, ecosystem preservation, and climate-change (Buda and DeWalle, 2002; DeWalle et al., 2000; Dow and DeWalle, 2000; Gellis et al., 2005; Huntington et al., 2004; Najjar, 1999; Reed et al., 2006). These broad challenges have actively engaged a growing array of individuals and organizations in the region's science and environmental policies to improve the Mid-Atlantic's sustainability. More than 700 groups or organizations are actively involved in advancing observation, prediction, and management efforts for the region's water resources. The technical goals for the MAWA portal are to (1) provide a meta-search resource for discovering and accessing the growing array of publicly available data for the Mid-Atlantic and (2) facilitate the emergent identification of the diverse stakeholder groups and projects actively shaping water resources planning and management within the region.

2. MAWA's software design & services-based framework

As illustrated in Fig. 2, the MAWA portal has been designed so that new data resources, web services, and enhanced data processing/management features can be added over the long-term evolution of the system. The MAWA portal's diverse and growing array of data sources includes more than 2500 regional gauging stations, 13,415 high-resolution image tiles, and a direct download for the full regional CONUS soils database (see http://www.soilinfo.psu.edu for details). MAWA exploits Microsoft's Virtual Earth mapping technologies to provide highly interactive searching,
visualization, and sharing functions that allow users to rapidly identify and analyze a broad range of geographically referenced data. The MAWA portal builds on the recent national hydrologic science observation and information management initiatives organized by the US Consortium of Universities for the Advancement of Hydrologic Science, Inc. (CUAHSI, http://www.cuahsi.org). More specifically, MAWA is a regional user of CUAHSI’s national hydrologic information system (Beran and Piasecki, 2009; Goodall

Fig. 1. Mid-Atlantic regional boundaries in the northeastern United States.

Fig. 2. MAWA mashes-up data and information from a variety of web-accessible sources and presents them in an interactive map-based web application to facilitate discovery within a geographic context.
et al., 2008; Horsburgh et al., 2009) and augments the system with a range of new features.

Fig. 2 provides an overview of the data management and services-oriented architecture underlying the MAWA portal. Relative to other tools that exist (e.g., the CUAHSI HIS, HydroSeek, Google Maps, etc.), MAWA provides a unique regional focus on providing data access and supporting peer-to-peer services. As shown in Fig. 2, the underlying information management in the system is meant to be extensible to a wide range of regionally relevant data resources and to provide other systems or portals with services-oriented access to MAWA hosted data. The overall user interface has been designed to maximize active view exploration of the region’s data resources using Adobe System’s Flex 3 SDK web application development platform [http://www.adobe.com/products/flex/]. The MAWA portal has focused primarily on a data services-based use of the HydroSeek [http://www.hydroseek.org (Beran and Piasecki, 2009)] and Pennsylvania aerial imagery repositories. MAWA extends these systems to provide highly interactive access, visualization, and download functionality for users. A unique feature of MAWA is the download access and public use of imagery at high resolution. This feature takes advantage of the PA MAP repository hosted by the Pennsylvania State University. The Flex-based services framework is easily extensible to new data sources and provides the potential for future extensions of the MAWA portal to support model prediction services (Athanasiadis, 2007; Mineter et al., 2003; Villa et al., 2009). Beyond services-based data identification and analysis, MAWA also seeks to encourage the emergent identification of the diverse groups and projects within the Mid-Atlantic region. The community building component of the portal uses a peer-to-peer strategy to allow users to geospatially tag the MAWA Virtual Earth exploration maps and provide summaries with links to their own sites. These Community Highlights are then added to the MAWA database and its search components.

3. A demonstration of MAWA search and share features

Fig. 3 illustrates a data search within the Upper Juniata watershed that demonstrates MAWA’s ability to identify community highlights and gauging in the watershed, analyze time series from the gauges, and download aerial imagery. This section of text will focus on guiding a replication of the search results shown in Fig. 3 (readers are encouraged to access MAWA before proceeding, http://www.mawa.psu.edu). Initially, by typing “Juniata” in the watershed box and selecting “Upper Juniata” the Upper Juniata’s watershed boundaries as defined by the United States Geological Survey’s (USGS) National Hydrography Dataset (see http://nhd.usgs.gov/index.html) are highlighted. Selecting the “Search” option at the top of the MAWA page then allows users to choose the “Dates” and “Geographic Area” for their subsequent search. In this example, clicking on “Geographic Area” will allow the user to focus geospatial search on solely the Upper Juniata by clicking “Use Watershed Area”. The triangular marker next to “Search Datasets” then allows the user to choose their datasets of interest.

Currently, MAWA has used four datasets to demonstrate its potential to support discovery and sharing of information by its user community. The Community Highlights search feature is tied closely to the “Share” function of the site. The “Share” function allows users to supply their emails and tag locations where they want to share links or information on their watershed-related efforts. MAWA then emails users with a web-form that can be easily filled out and submitted to the system’s search database. As an example, clicking on the “Community Highlights” option reveals alternative search strategies. By first selecting the “include in search” box, the community highlights will become active in the multi-data MAWA search engine. In this example, typing “Tang” under the Last Name box will identify a study that was conducted in the Spruce Creek sub-watershed of the Upper Juniata. Users can tag multiple entries and websites based on their

![Fig. 3. Example screen shot from a search in the Upper Juniata watershed. The results include both the HydroSeek Spruce Creek stream gauge hydrograph from 2006 to 2008 and the potential to download the aerial image denoted in the red bounding box.](image-url)
The next dataset, RTHnet, designates the Real-Time Hydrologic Monitoring Network research watershed. The purpose of RTHnet’s inclusion is to demonstrate how different real-time experiments can potentially exploit MAWA’s data discovery and share features. RTHnet is a university-based research effort that provides unique high-resolution time series for a wide range of measurements. The RTHnet system provides access to a range of streamflow, groundwater, micrometeorological, and soil moisture observations. Moreover, the RTHnet system will evolve to aid in publishing the diverse data that are being collected in the region’s Shale Hills Critical Zone Observatory [Brantley et al., 2006; http://www.czen.org/]. The Juniata/Shale Hills critical zone observatory has synergistic field experiments that when taken as a whole, seek to better understand the dynamic couplings and feedbacks between the fluid, mineral, gaseous, and biotic components of the Earth’s terrestrial system from the outer envelope of vegetation down to the lower limits of groundwater. In this example, again selecting “Include in Search” and clicking the Keyword box yields a range of data to search (e.g., “air temperature”).

The next data source is a services-based data search that draws on the HydroSeek (Beran and Piasecki, 2009) and the CUAHSI hydrologic information system (Goodall et al., 2008; Horsburgh et al., 2009) to identify a very broad range of time series on both water quantity and quality measurements from the USGS, the US Environmental Protection Agency, and the Chesapeake Bay Information Management systems. Building on the keyword ontology developed by Beran and Piasecki (2009), MAWA users can click HydroSeek’s “include in search” option and then type a key term to search for discovering regional data. In this example, typing “streamflow” will identify the gauging data throughout the Upper Juniata. Our goal in this particular example will be to quickly find and analyze 2 recent years of streamflow data in the Spruce Creek sub-watershed. Spruce Creek is a pristine headwater watershed within the region that has very significant ecological and recreational value.

The final dataset available in MAWA allows users to identify and download high-resolution orthophotography throughout the Mid-Atlantic region. Clicking the “Aerial Imagery” dataset and “include in search” will then identify the available regional photography in and around the boundaries of the Upper Juniata. To summarize to this point, the geographic area, the data date range, and the specific datasets to be included in the Upper Juniata search have all been specified. Pressing “Search” activates the collective search of community highlights for “Tang”, the RTHnet data for “air temperature”, the HydroSeek data “streamflow”, and includes a call for the Upper Juniata’s aerial imagery. Successful search will launch the Results menu where by clicking on each dataset’s name, the user can identify how many records have been discovered (e.g., 1 highlight, 2 RTHnet records, 14 HydroSeek stations, and 638 image tiles). For the community highlights, RTHnet, and HydroSeek data highlight, 2 RTHnet records, 14 HydroSeek stations, and 638 image tiles). For the community highlights, RTHnet, and HydroSeek data sources users can choose the “Add to Map” option to put tags on the Virtual Earth map that guide interaction and exploration. Alternatively, these datasets also have a “Show Table” feature that will allow users to explore the data in a tabular format.

Adding the Community Highlight to the map will reveal a study by Tang analyzing flood forecasting in the Spruce Creek watershed and provide a link to a journal abstract describing the work in more detail. The Community Highlights provide a brief but potentially powerful geospatial tag-and-search capability for highlighting regional efforts. As shown on Fig. 3, adding the 14 streamflow gauging stations from HydroSeek yields a spatial tag very near the Tang community highlight. Although the Aerial Imagery can be added to the map, this is a very large dataset and a time intensive exploration mode. Alternatively, users can sweep their cursor over the imagery tile numbers in the Search Results panel to display the boundary of each tile. The scroll bar can be used to quickly cycle through the image tiles until the boundaries within the area of interest are displayed. In Fig. 3 example, the tile boundary titled “47001850PAS” near the Spruce Creek HydroSeek and Community Highlight results is displayed in red and can be downloaded by the user. This zoomed view was attained by clicking and holding the left mouse button while positioning the center of the viewing frame on the data tags. Using the mouse’s scroll wheel it is possible to zoom in and out of the Spruce Creek streamflow gauge.

Beyond the imagery, MAWA also allows users to “Graph/Download” HydroSeek data (assuming the source service feed is reliable). In this example, right-clicking on the Spruce Creek gauging station yields a menu option to “Graph/Download” that when selected opens an interactive data exploration tool. Selecting the “Variables” control displays data available at the gauging station and left-clicking the “Create Graph” button generates the hydrograph shown in Fig. 3. Alternatively, the user can download the data directly by left-clicking the “Download” button. Although this is a simple example, MAWA contributes a rapid mechanism for sharing and communicating information within the Mid-Atlantic.

4. Summary

Despite broad interest in the watersheds across the Mid-Atlantic region of the US, data access, publication, and sharing has largely been segmented across major governmental agencies. The tremendous data repositories at both the federal and state levels in the Mid-Atlantic require a high degree of user sophistication and often internal organizational knowledge for access and use. This paper demonstrates how the MAWA initiative uses emerging innovations in services-based information management and publication (Beran and Piasecki, 2009; Goodall et al., 2008; Horsburgh et al., 2009) to facilitate broad community engagement in watershed science and management across the Mid-Atlantic of the US. Readers interested in more information on the use of the MAWA system are encouraged to view the multi-media video tutorials in the Help menu.

References


ARTICLE IN PRESS


