Design and Development of Web Services for Accessing Free Hydrological Data from the Czech Republic

Jiří Kadlec[1], Daniel P. Ames[1],

[1]Idaho State University, 995 University Blvd, Idaho Falls, ID, 83402, USA
Kadljiri@isu.edu,dan.ames@isu.edu

Abstract. The target of the open source hydrodata.cz web application and web services is to bridge the gap between modelers and field hydrologists by providing free, publicly accessible precipitation, snow, temperature, and discharge data series from multiple organizations in the Czech Republic using the standardized the CUAHSI WaterML data format and search services. Data providers contributing to this effort include the CUAHSI, NOAA, Charles University experimental watershed network, Czech watershed authorities, and networks of volunteer observers. Time period covered includes the floods of 2006, 2009, 2010 and 2011 as well as the 2007 drought. Adopting the WaterML standard also allows the use of other free third party client side web based and desktop data discovery software tools – which has the added benefit of making the Czech Republic hydrological data more readily accessible to the international community.

Keywords: Hydrology, web services, interoperability, information system

1   Introduction

Hydrologic simulation models are used extensively by scientists and water management authorities for evaluating the response of water ecosystems to land use and climate changes. A large number of open source simulation models are available, including SWAT [1],WASIM-ETH [2], HSPF [3] and AGNPS [4]. Using open source tools for rainfall-runoff simulation (as opposed to proprietary software) can give the researcher better understanding of the approximations used in the model because of access to source code and hence inherent model assumptions and implementation. It also encourages improvement of the model by the user community [5]. The obstacle in implementing either open source or proprietary models by researchers in the Czech Republic is limited availability of precipitation, discharge and soil data for model parameter setup, calibration and validation. Similar to other European countries, the Czech Republic has a long record of systematic meteorological observation and soil mapping. These observations are rarely accessible to the general public. Fig. 1 shows an example of obtaining hourly precipitation data for a recent flash flood event for a small catchment in south-west Bohemia. The process includes contacting several organizations, finding the responsible data managers, signing the data agreement, and

transferring the data fee. In each organization contacted by the user, the data manager must connect to an internal database, find any available observation sites, export database query results to text file and deliver the text file to the user.
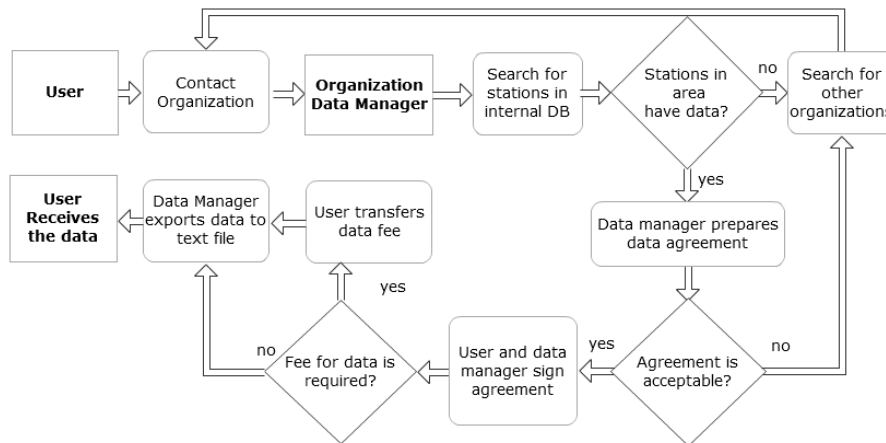

Fig. 1. Obtaining Precipitation data from Blšanka watershed, Czech Republic

Some free data sources already exist in the Czech Republic. Indeed, as required by the information access law of the Czech Republic and emergency early warning system, many government organizations are required by law to publish real-time records of streamflow on the voda.gov.cz [6] internet server. A user with programming experience may set up a method of automated, scheduled data retrieval from this server (Fig. 2). This approach has already been used by the raft.cz white water rafting information web site [7] for the Czech rivers and streams to determine the best time of the year for river navigation. A problem with this approach is large amount of HTML parsing code that requires ongoing maintenance when the structure of the original HTML page changes.

Fig. 2. Getting stream stage data – automated method used by raft.cz

1.1   CUAHSI Hydrologic Information System

The Consortium of Universities for Advanced Hydrologic Science Hydrologic Information System (CUAHSI HIS) is a distributed internet system for sharing hydrologic time series field and derived observations from international and U.S

agencies, universities, experimental watersheds and other organizations [8]. The system consists of three main components: *HydroServer*, *HydroCatalog* and *HydroDesktop* (Fig. 3). This configuration follows the well-established "publish-find-bind" model used throughout the Internet for web services [9].
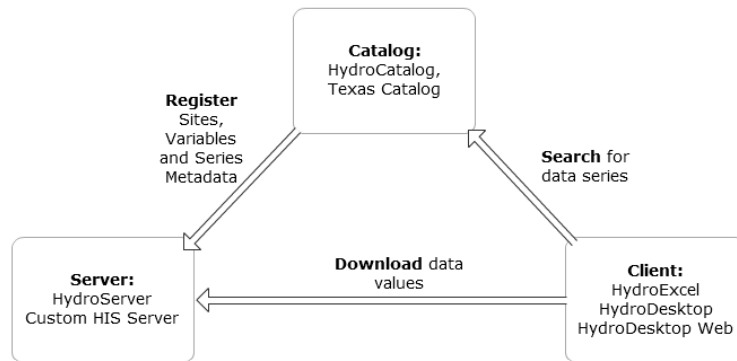

Fig. 3. Main components of the CUAHSI hydrologic information system

1.2 HydroServer

A "HydroServer" is any server connected to the Internet which exposes a standard web service defined by CUAHSI [10]. Primary web methods of the web service include GetSites, GetSiteInfo and GetValues. The GetSites method returns all sites with available data. GetSiteInfo returns the list of variables available for a specific site. GetValues returns the time series for the specified site, variable and time interval. The results of the web methods are returned in the standardized WaterML format [11]. WaterML is an extended variant of the XML language and includes specific elements for defining attributes of hydrologic time series. WaterML requires the data publisher to include complete metadata about the time series. Each HydroServer web service method response is a fully self-describing WaterML document which contains metadata for units, data source, method, quality control level, site location and variable data type. CUAHSI is working together with the Open GIS Consortium (OGC) and World Meteorological Organization (WMO) to estabilish WaterML 2.0 which will be fully compliant with the widely used WFS (Web Feature Service) and SOS (Sensor Observation Service) standards. The adoption of services oriented architecture exposes a platform – independent interface which may be implemented by any server based web application technology. A default implementation (HydroServer) is provided by CUAHSI as open source software [12]. The core of the HydroServer is a Microsoft SQL Server database using the Observations Data Model

(ODM) schema [13]. The comprehensive schema of the ODM database is designed for efficiently storing climate, hydrology and water quality time series observations. The official release of CUAHSI HydroServer requires a virtual private or dedicated server with the Microsoft Windows Server operating system. A simplified implementation of HydroServer on a commercial web hosting provider is described later in this paper.

1.3   HydroCatalog

HydroCatalog provides a centralized registry of publicly accessible HydroServers. It exposes a catalog web service with a set of web methods that allow searching for sites and time series across multiple servers. The main web method of the web service is *GetSeriesCatalogForBox*. Given a latitude and longitude bounding box, keyword, start time and end time, the metadata of all time series which match the search criteria are returned in XML format. In order to be searchable by HydroCatalog, the HydroServer must be registered with a central catalog. A key step in the registration process is the assignment of keywords. For each variable measured by the HydroServer, the matching keyword is chosen from the CUAHSI Ontology Keyword list [8]. After registration, HydroCatalog runs a periodic harvesting operation which ensures that any additional sites or variables may be searched. The CUAHSI catalog, HIS Central, is operated by the San Diego Supercomputer Center (SDSC).

1.4   Client Applications

Numerous desktop and web based client applications have been to support the discovery, download and analysis of hydrological time series data made available by the CUAHSI HIS. The desktop applications include HydroExcel and HydroDesktop. HydroExcel [8] supports the download and descriptive statistical analysis of time series from a user specified HydroServer web service. HydroDesktop is an open source geographic information system (GIS) – based tool with special features for hydrological data visualization and analysis [14]. The map component of HydroDesktop enables detailed spatial analysis of the observation sites. Recently a web based client similar to HydroDesktop has been developed [15] with the same core functionality as HydroDesktop (search, map, graph and data export). It is expected that additional client applications will be developed in the near future, customized for specific regions or research domains.

2   Software Design and Implementation

2.1   Analysis of Requirements

For a more detailed analysis of data and software requirements we have selected precipitation, air temperature, water stage, discharge and snow cover data

requirements. As a first step, the potential data providers with existing meteorological and hydrological observations covering the Czech Republic were examined.

2.2 Evaluation of Data Sources

Hourly **precipitation** time series are published by the Czech Hydrometeorological Institute (CHMI) flood warning service (HPPS) for previous 7 days from automated rain gauge stations. The CHMI radar network provides images of radar reflectivity for previous day in 2 km, 15 minute resolution. A series of combined 24-hour radar / ground station precipitation maps is also available starting January 2007. In these maps, precipitation for each pixel is represented showing discrete color interval. Some parts of the Czech Republic have coverage of radar networks in Germany, Poland and Slovakia. For the southern part of the country, the global satellite derived precipitation data products (PERSIANN) are available [16]. Precipitation is also published from automated sites of the Czech watershed authorities on the government water portal [6] and by university experimental watershed networks. Precipitation data are susceptible to large measurement errors. By combining multiple systems, automated quality control can be enhanced. **Air temperature** hourly time series are available from the Czech watershed authorities and from university experimental watersheds. The existing NCDC Hourly data web service in the CUAHSI HIS system also provides long term temperature time series from several locations in the Czech Republic. Water **stage** and **discharge** observations are publicly available on the government water portal [6]. Additional observations are published by municipality owned early warning systems [17] and by university experimental watershed networks. **Snow cover** is an important input parameter for rainfall-runoff modeling during the winter period. Daily observations of snow depth are publicly accessible on the web sites of CHMI and watershed authorities. Several volunteer groups also measure snow depth. This includes HS Brdy (Brdy mountain guard) whose cross country skier observers provide detailed coverage of snow conditions in Brdy, the largest forested mountain are in inland Bohemia [18]. Table 1 shows the list of data providers contributing to this system, including the current status.

Table 1. Organizations from the Czech Republic contributing to CUAHSI web services

Organization	Type of observations
CHMI	Snow depth, precipitation, stage, discharge
Povodí Ohře	Snow depth, precipitation, air temperature, stage, discharge
Povodí Vltavy	Precipitation, stage, discharge
Charles University Prague	Water stage, precipitation, temperature, solar radiation, conductivity, dissolved oxygen
HS Brdy	Snow depth

2.3   Implementation of HydroServer for Small Organizations

For smaller data providers, including volunteer observer groups, the main obstacle for providing data in the form of HydroServer web services are associated with hardware and software requirements of the default CUAHSI HydroServer software. These organizations frequently use third-party webhosting services. However, the original HydroServer software was designed to work on large dedicated university servers. Initial testing has shown that third party webhosting providers do not allow running applications under full trust and restrict the usage of some software libraries which were required by the original HydroServer. To address such limitations, the source code of HydroServer was modified, creating a customized product: HydroServer Lite specifically to meet the needs of a Czech HIS implementation. HydroServer Lite may be deployed on any ASP.NET webhosting service, reducing maintenance costs for the organization.

2.4   Overcoming Data Security Concerns

The large Czech government supported organizations (CHMI, watershed authorities) were initially reluctant to make any changes on their servers in support of the hydrodata.cz web services and the CUAHSI HIS system. As a result, an alternative strategy has been implemented (Fig. 4). The hydrodata.cz web server has been set up on a large third party webhosting service. Each day, an automatic scheduled task is launched. This program sends a request to the CHMI web pages, parses the HTML code using regular expression template, extracts time series observations and stores these to the hydrodata.cz database. The standard HydroServer web service is built on top of this database. In case the formatting of the CHMI web page changes significantly, a message is sent by the system, notifying the hydrodata.cz manager that a change in the html parser template is required. With increasing future awareness of the CUAHSI HIS system and more emerging applications, it is expected that the CHMI and other organizations will gradually implement the HydroServer web services directly on their own servers as HIS becomes more widely adopted.
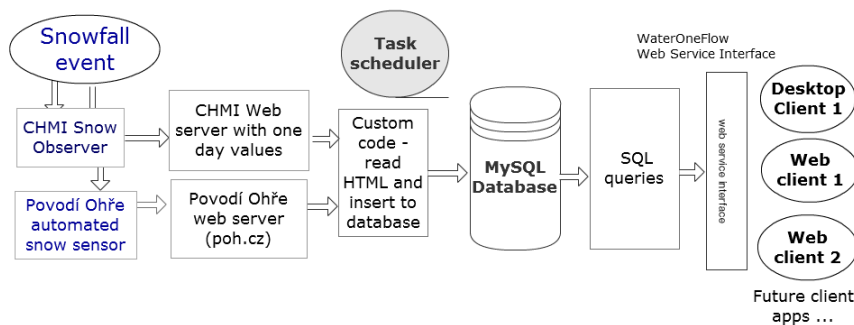

Fig. 4. Adaptation of HydroServer for Czech Republic Snow HydroServer web service

3   Conclusion

The Czech Republic has become one of first regions in the world outside North America full, countrywide coverage of snow and precipitation data made publicly accessible through the CUAHSI HIS (Fig. 5). After fully integrating discharge records, this will allow to re-calculate the surface water balance of any watershed in the country. Other potential applications include hydrology education and agricultural management and planning. The hydrodata.cz web services and application are in the prototype testing stage. By developing the services as an open source system, the authors are convinced that contributions from new users will lead to improvement in system usability. Major developments planned in near future include: (1) Integration of satellite and radar estimates of precipitation, which requires more efficient handling of gridded datasets, (2) Integration of downscaled global climate model outputs, (3) Addition of publicly accessible services providing soil and land use maps using the WMS / WCS standards, (3) Inclusion of more volunteer meteorological observer networks in the system, and (4) Implementation of pay-per-access method to enable access to data sets from institutions which require the purchase fee. By implementing an automated internet payment system, more users will be able to promptly access data sets which require additional financial support. Adopting the WaterML standard also allows the use of other free third party client side web based and desktop data discovery software tools – which has the added benefit of making the Czech Republic hydrological data more readily accessible to the international community.
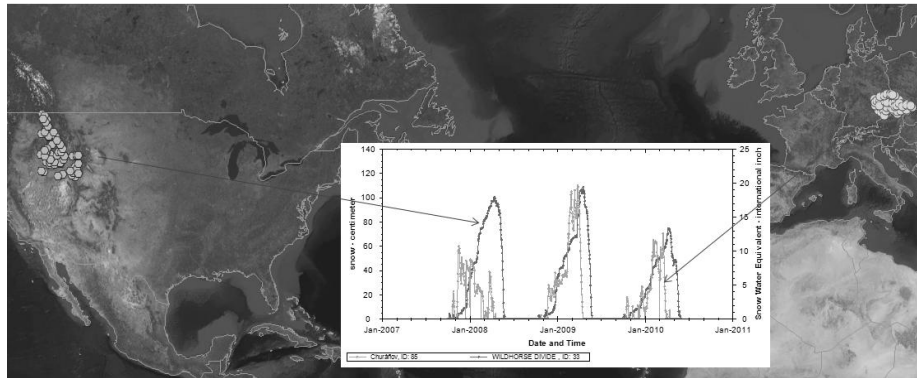

Fig. 5. Map of snow cover sites in Idaho Rocky mountains (USA) and Czech Republic with a comparison of snow depth time series graph in HydroDesktop

References

1. Arnold, J. G., Srinavasan, R., Muttiah. S., Williams, J.R: Large Area Hydrologic Modelling and Assessment Part I: Model Development. J Am Water Resour As, 34, 73–89. (1998)

2. Schulla, J., Jasper, K..: Model Description WaSiM-ETH. IAC ETH Zürich, pp. 166. (2001)

3. Donigian, A.S., Bicknell, B.R., Imhoff, J.C.: Hydrological Simulation Program – Fortran (HSPF). In: Singh, V.P.(ed.). Computer models of watershed hydrology, pp. 395-442. (1995)

4. Young, R.A., Onstad, C.A., Bosch, D.D., Anderson, W.P.: AGNPS: A nonpoint-source pollution model for evaluating agricultural watersheds, J Soil Water Conserv 44(2), 168-173. (1989)

5. Gregersen, J.B., Gijsbers, P.J.A., Westen, S.J.P.: OpenMI: open modelling interface, J Hydroinform 9, 175–191. (2007)

6. Water Management Information Portal, http://voda.gov.cz

7. Raft.cz Water Sport Portal, http://rivers.raft.cz

8. Maidment, D.R.: Bringing water data together, J Water Res Pl 134(2), 95–96. (2008).

9. OGC Reference Model. Open Geospatial Consortium, 2008-11-11. OGC 08-062r4. http://www.opengeospatial.org/standards/orm

10. Tarboton, D.G., Horsburgh, J.S., Schreuders, K., Maidment, D.R., Zaslavsky, I., Valentine, D.W.: The HydroServer Platform for Sharing Hydrologic Data, American Geophysical Union, fall Meeting 2010, abstract #H53H-03 (2010)

11. Zaslavsky, I., Valentine, D., Whiteaker, T.: CUAHSI WaterML. OGC 07-041r1, Open Geospatial Consortium Discussion Paper (2007)

12. Horsburgh, J. S., D. G. Tarboton, M. Piasecki, D. R. Maidment, I. Zaslavsky, Valentine, D., Whitenack, T.: An integrated system for publishing environmental observations data, Environ Modell Softw, 24, 879-888 (2009)

13. Horsburg, J., Tarboton, D.G., Maidment, D.R., Zaslavsky, I.: A relational model for environmental and water resources data, Water Resour. Res, 44 (2008)

14. Ames, D. P., Horsburgh, J., Goodall, J., Whiteaker, T., Tarboton, D., Maidment, D.: Introducing the Open Source CUAHSI Hydrologic Information System DesktopApplication (HIS Desktop), 18th World IMACS/MODSIM Congress, Australia (2009)

15. HydroDesktop Web Map Application, http://www.hydromap.info

16. Hsu, K.L., Gao, X.G., Sorooshian, S., Gupta, H.V., 1997. Precipitation estimation from remotely sensed information using artificial neural networks. J Appl Meteorol 36 (9), 1176–1190 (1997)

17. Hladiny.cz Early Warning System, http://www.hladiny.cz

18. HS Brdy (Brdy mountain guard), http://www.hsbrdy.com