Web Based Access to Water Related Data Using OGC WaterML 2.0

Adrian Almoradie

Dept. of Integrated Water Systems and Governance
UNESCO-IHE Institute for Water Education
Delft, the Netherlands
a.almoradie@unesco-ihe.org

Andreja Jonoski

Dept. of Integrated Water Systems and Governance
UNESCO-IHE Institute for Water Education
Delft, the Netherlands
a.jonoski@unesco-ihe.org

Ioana Popescu

Dept. of Integrated Water Systems and Governance
UNESCO-IHE Institute for Water Education
Delft, the Netherlands
i.popescu@unesco-ihe.org

Dimitri Solomatine

Dept. of Integrated Water Systems and Governance
UNESCO-IHE Institute for Water Education
Delft, the Netherlands
d.solomatine@unesco-ihe.org

Abstract—As the Open Geospatial Consortium (OGC) has adopted WaterML 2.0 as encoding standard for representing hydro-meteorological time series data, the water community is in need of tools and methods for delivering such data over the web.

This article presents experiences with one approach for publishing water-related data over the web based on the OGC WaterML 2.0-GeoServer framework and methods developed by Australia's Commonwealth Scientific and Industrial Research Organisation (CSIRO).

This implementation is one component of a web based flood information system for Somes Mare basin in Romania, which has been developed within the enviroGRIDS EU FP7 research project.

Keywords—WaterML 2.0; GeoServer; Flood information dissemination; web-based; Somes Mare

I. INTRODUCTION

Environmental disasters such as floods are expected to increase because of rapid urban development, population growth and climate change. Such example is the flood-related disaster that happened in Pakistan in the year 2010. The flood affected over twenty million people, more than the combined major disaster of the 2005 Pakistan Earthquake, 2005 USA Katrina Cyclone, 2008 Myanmar Nargis Cyclone, 2004 Indian Ocean Basin Tsunami and the 2010 Haiti Earthquake.

In Europe, floods are the most common cause of natural disasters. During the summer of 2013 some parts of Germany and Czech Republic experienced severe flooding. Record breaking water levels, were observed in 2013, on the Elbe and Danube rivers. Estimated damages in Germany alone may have reached $12 billion.

As flood related disaster increases, it is likely that decision makers and authorities will increase their hydrological monitoring for the purpose of better management and planning of flood risk.

Increased monitoring will produce large amounts of data that needs to be properly managed and utilized in different modeling and decision support applications. Further, to better make use of available data coming from different sources (i.e. institutions, organizations and flood management authorities) there is a need for web-based systems for sharing and accessing these data.

There are several existing data publication methods [1] used by different organizations. In general these organizations have different systems in collecting, formatting, archiving and publishing data.

Since environmental data may come from different sources, these data are likely to be syntactically and semantically heterogeneous. Syntactic heterogeneity means different data structure or format [1]. Semantic heterogeneity is defined as the differences in the objects and attributes that define the data, leading to disagreement on the meaning, interpretation and use of the same data [2]. Semantic heterogeneity can further be divided in two types, structural and contextual, as further elaborated in [1].

To overcome data heterogeneity and to better manage, share and analyze these large amounts of data there is a need to make use of standards (recognized by scientific community) that help these data to be organized and published [1]. None of the existing data publication methods has been widely embraced by the scientific community as a standard for publishing data. However there has been relevant progress when the Open Geospatial Consortium (OGC) community started the collaborative effort on developing such standards for geospatial data, many of which are of high relevant to environmental applications. In the area of water, the OGC recently accepted the Water Mark-up Language (WaterML) 2.0 as a standardized marked up language for publishing water-related time series data.

Flood risk management (FRM) will greatly benefit in using standards for publishing, sharing and accessing data. However,

how to properly make use and present these data to users currently presents a major challenge.

Users can be decision makers, experts, flood authorities, stakeholders and citizens. An essential component for flood risk management is raising awareness of potentially-affected citizens. However for the past two decades environmental managers and authorities have seen the growing need for people to more actively participate in the environmental management [3]. Incorporating stakeholder's beliefs, values and their local knowledge of the environment in FRM will lead to more sustainable measures and decisions [4-5].

Web-based systems are increasingly seen as potential tools for flood information sharing, dissemination and participation, which may greatly enhance flood risk management. Moreover, the use of standards for publishing data on these web-based systems will enable a more efficient data transmission, subsequent analyses and decision making. These tasks are performed by different users and require different applications for meeting their needs. The integration of such diverse application can greatly benefit from the use of established standards for data publishing and sharing.

This article presents experiences for publishing water-related data over the web based on the OGC WaterML 2.0-GeoServer framework and methods developed by Australia's Commonwealth Scientific and Industrial Research Organization (CSIRO).

This implementation is one component of a web-based flood information system (FIS) developed for Somes Mare basin in Romania. The FIS was developed to be used by flood management authorities and potentially-affected citizens. The latest technologies for collection, archiving and sharing of environmental data, using web-based Spatial Data Infrastructure (SDI) were implemented.

This system was developed within a research project entitled enviroGRIDS at the Black Sea Catchment, funded by the EU FP7 Research Framework.

The article is organized as follows: Section II presents WaterML 2.0 as a standardized markup language. WaterML2.0-GeoServer architecture is presented in section III. Overview of the Somes Mare FIS development and implementation of the WaterML 2.0-GeoServer is presented in section IV. The final section discusses the advantages of using this approach, experiences and future development.

II. WATERML 2.0 AS A STANDARDIZED MARKUP LANGUAGE

In the past decades there have been initiatives from scientific communities to make use of standardized markup languages to address data heterogeneity from different sources. Such examples are the Earth science Markup Language (ESML) [6], Ecological Metadata Language (EML) [13], Observations and Measurements (O&M) by OGC [7] and Water Markup Language (WaterML) [8].These standardized markup languages are presented in Extensible Mark-up Language (XML) format.

Their differences are the vocabularies used (e.g. discharge-streamflow, precipitation-rainfall) and how the objects and their attributes that describe the data are structured.

Different versions of WaterML have already been used by several main international organizations, such as Consortium of Universities for the Advancement of Hydrologic Science, Inc. (CUAHSI) from USA, Australia's Commonwealth Scientific and Industrial Research Organization (CSIRO) and Bureau of Meteorology (BOM), United States Geological Surveys (USGS) and National Oceanic and Atmospheric Administration (NOAA) from USA. Although these organizations, institutes and scientific communities have already used this language, one issue that remains is its interoperability with other systems. To address this issue, an international group of organizations that encourage development of open standards (OGC) accepted and endorsed WaterML 2.0 schema [9] as an encoding standard for publishing time series of hydrological observation data. WaterML 2.0 is an updated version of WaterML that incorporates the OGC O&M standards.

In Europe, scientific communities, organizations and institutes are encouraged to make use of this standard (WaterML 2.0) and establish a system for publishing hydrological observation data in WaterML 2.0 format.

III. WATERML 2.0-GEOSERVER ARCHITECTURE

One known system that publishes time series in WaterML via web services is the CUAHSI-Hydrologic Information System (HIS). CUAHSI-HIS web services system called WaterOneFlow has been tested and is in use by several US environmental agencies such as USGS, NOAA and NASA.

Although CUAHSI-HIS is free to install, it requires several commercial products for it to be functional, such as Microsoft Windows Server, Microsoft ASP.NET 2.0, Microsoft SQL Server 2008, ESRI ArcGIS 9.3.1 Desktop and Server for .Net (Enterprise Advanced).

As these commercial software packages are expensive to acquire and maintain, scientific communities, organizations and institutes are searching for alternatives in publishing data over the web. Such alternative is to make use of an open source or freely available systems for publishing data over the web.

A team from Commonwealth Scientific and Industrial Research Organization (CSIRO) in Australia developed a framework and methods that implements WaterML 2.0 schema using the GeoServer Web Feature Services (WFS). CSIRO developed the WaterML 2.0-GeoServer to publish water storage time series information (from lakes and reservoirs) from the Australian Bureau of Meteorology (BoM) [10].

GeoServer and its components are an open source - general public licensed (GPL) software/technology (e.g. PostgeSQL DB, Tomcat server, etc.). GeoServer implements OGC standards for publishing spatial data. Such standards are the Web Mapping Services (WMS), Web Feature Services (WFS) and Web Mapping Content (WMC).

Fig. 1 presents the WaterML 2.0-GeoServer architecture. Aspects of data storage, discovery and access, and those related to GeoServer configuration are briefly presented below.
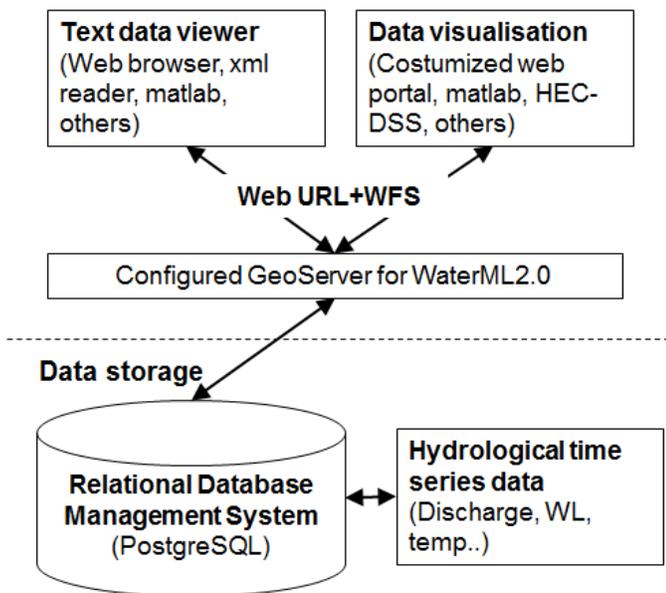

Fig. 1.   WaterML 2.0-GeoServer Architecture

A. Data storage, discovery and access

Discovery and access to data is made possible through the use of Universal Resource Locator (URL) that host the GeoServer web services accompanied with WFS operators. With WFS it is possible for clients to query the data structure and the actual data. WFS can perform several operations to query the data. Such operators are the *GetCapabilities* - retrieves a list of the server's data; *DescribeFeatureType* - retrieves information and attributes about a particular dataset; and *GetFeature* - retrieves the actual data.

For WaterML 2.0 the WFS GetFearture is the main operator key to retrieve and query the actual data. For details on querying keys for WFS please see http://docs.geoserver.org/stable/en/user/services/wfs/.

Calling the URL with WFS operators can be done via web browsers or other tools that interpret an XML format. There are also some software tools that can read directly WaterML 2.0 schema and present the data in a table or charts, such as the HEC-DSS software tool by the US Army Corps of Engineers. HEC-DSS is a database management system for HEC modeling softwares (e.g. HEC-RAS and HEC-HMS).

For data storage and management the framework made use of an open source - GPL relational database management system technology (postgreSQL) compatible with GeoServer.

B. WaterML 2.0 - GeoServer configuration

The GeoServer software was re-configured by installing plug-ins for database connection, updating the schema files and GeoServer properties (e.g. workspaces). The updated Schema files contain the WaterML 2.0 structure.

The WaterML 2.0-GeoServer framework structured the information about the data in four main parts: (1) time series data, (2) geographical information and geometry, (3) data provider details and (4) details about the stations.

In reference to CSIRO BoM's implementation of WaterML 2.0-GeoServer for water storage observation [10], Table I presents the schema files description and its corresponding database and Fig. 2 presents the schematic of database table relationships. As will be shown later this structure has been adapted for the implementation of the application presented in this article.

TABLE I.   CSIRO BoM IMPLEMENTATION OF WATERML 2.0-GEOSERVER: SCHEMA XML WITH ITS CORRESPONDING DATABASE TABLE

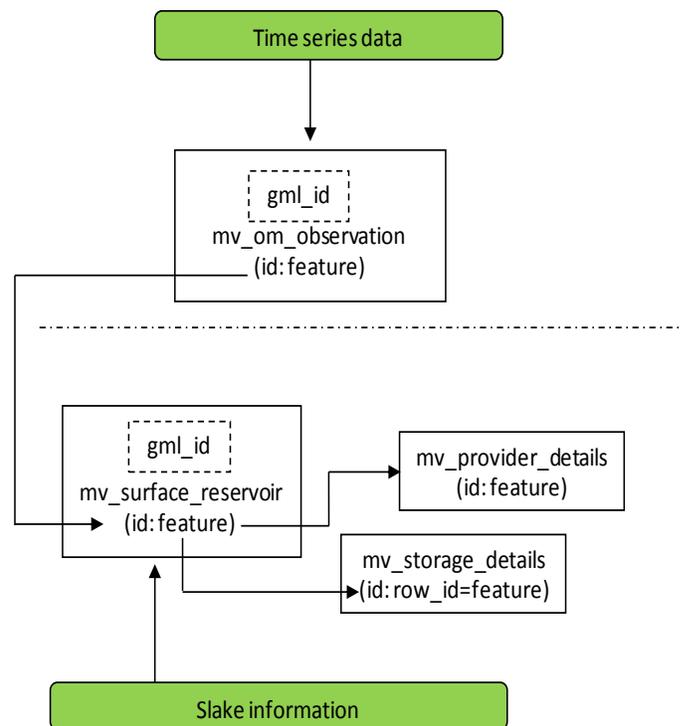| Workspace	Xml schema	Slake database table	Description
om	OM_Observation.xml	mv_om_observation	time series data
slake	SurfaceReservoir.xml	mv_surface_reservoir	lake information and geometry
	ProviderDetails.xml	mv_provider_details	data provider details
	StorageDetails.xml	mv_storage_details	lake storage details

Fig. 2.   Database table relationships

Reference [10] provides a development guide (designed for BoM water storage observation) to configure GeoServer for WaterML 2.0.

IV.   FIS IMPLEMENTATION OF WATERML 2.0-GEOSERVER

This section first presents an overview of the web-based FIS followed by the implementation of WaterML2.0-GeoServer.

A. Overview of the Somes Mare Flood Information System

The web-based FIS was designed and developed for flood information dissemination and stakeholders-citizen participation. Latest web development technology freely available was used in its development. The following briefly presents the case study area and the FIS conceptual design.

1) Case study area

Somes Mare is a catchment of the larger Somes basin. The basin is located in the northern part of Romania. Somes Mare has an area of about 5078 km^2. The catchment is vulnerable to flooding especially during the spring season when snow from the mountains melts. Flood is most devastating when combination of rainfall with snowmelt occurs. In the past 50 years the most important flood was the one of 1970, corresponding to the 100 years return flood.

Lately, on Somes Mare River there are many occurrences of flash floods. The most devastating one was the one from 2009. This situation demonstrates the need for further studies in order to build and implement a better flood risk management strategy in the Somes Mare catchment.

Flood risk awareness of the citizens and information sharing is one of the approaches in management of floods. An innovative solution to reach the citizens and share information is through a web-based flood information system.

2) Conceptual design

The web-based FIS was designed to be simple, informative, interactive, customizable and flexible. Map based applications were extensively used for publishing geospatial data and accessing information. Furthermore, web infrastructure services and data standards were used.

The FIS has three main components: (1) FRM awareness, (2) Flood information access and (3) Citizens participation. Fig. 3 presents the conceptual design of the FIS portal.

The component FRM awareness is intended to raise citizens' awareness on the catchments flooding problems and its management plans. Flood information access is intended to raise the citizens' and stakeholders awareness on local flooding. Historical floods, observed data and model results (time series and flood maps) are presented. Moreover, flood information access has a sub-component for data access. This may be of less interest to users such as citizens and non-expert stakeholders, but more for professional users. Through this sub-component users can access the actual data (hydro-meteorological, time series and spatial data). Flood information access is provided by using the OGC standards such as WMS, WFS and WaterML 2.0. Details on the implementation of the OGC WaterML 2.0-GeoServer standards in FIS are presented in the later section.

The citizens' participation component provides the citizens and stakeholders with opportunities to discuss flood related issues, share information and timely report on local flooding.

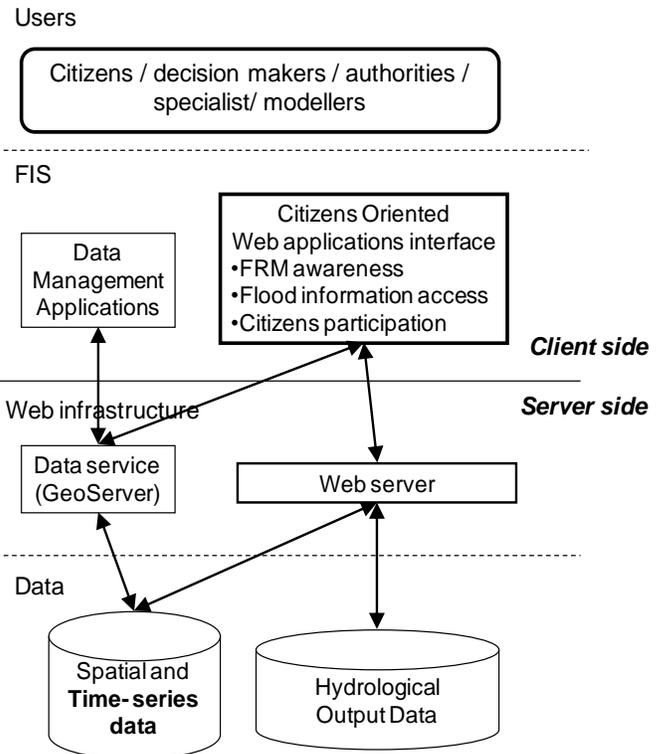The FIS portal can be accessed through the following URL:

- http://hikm.ihe.nl/envirogrids/Platform/Somes/

Fig. 3.   Web-based FIS generic conceptual design

Reference [11-12] presents more details on the FIS design, implementation and its evaluation.

B. Implementation of WaterML 2.0- GeoServer

The web-based Flood Information System (FIS) for Somes Mare basin of Romania implemented the WaterML 2.0-GeoServer framework and methods developed by CSIRO to publish hydro meteorological time series.

The original version of GeoServer was re-configured by installing plug-ins and updating the schema files and properties (updated schema files were provided by CSIRO). To have the correct database structure the database "Slake" from CSIRO was uploaded and updated with the id's, data and related information of the Somes Mare basin. Table II presents the data tables that were updated and renamed. Fig. 4 is a sample snapshot of the modified Id.

TABLE II.   DATA TABLES MODIFICATION

Slake database table	Somes Mare database table
mv_om_observation	mv_om_observation
mv_surface_reservoir	*mv_hydromet*
mv_om_observation	mv_om_observation
mv_storage_details	*mv_hydrometdetails*

SLAKE		Somes Mare

gml:id="slake.surfacereservoir.10.observation.67"; | gml:id="sbasin.hydrometeo.1.observation.1"

Fig. 4. Schema-database modification of id.

As previously mentioned, one sub-component of the Somes Mare FIS portal is the "Data access" (Fig. 5). The data access component allows users to view and download available spatial and time series data. The Water ML 2.0 formatted time series data are accessible through this sub-component.
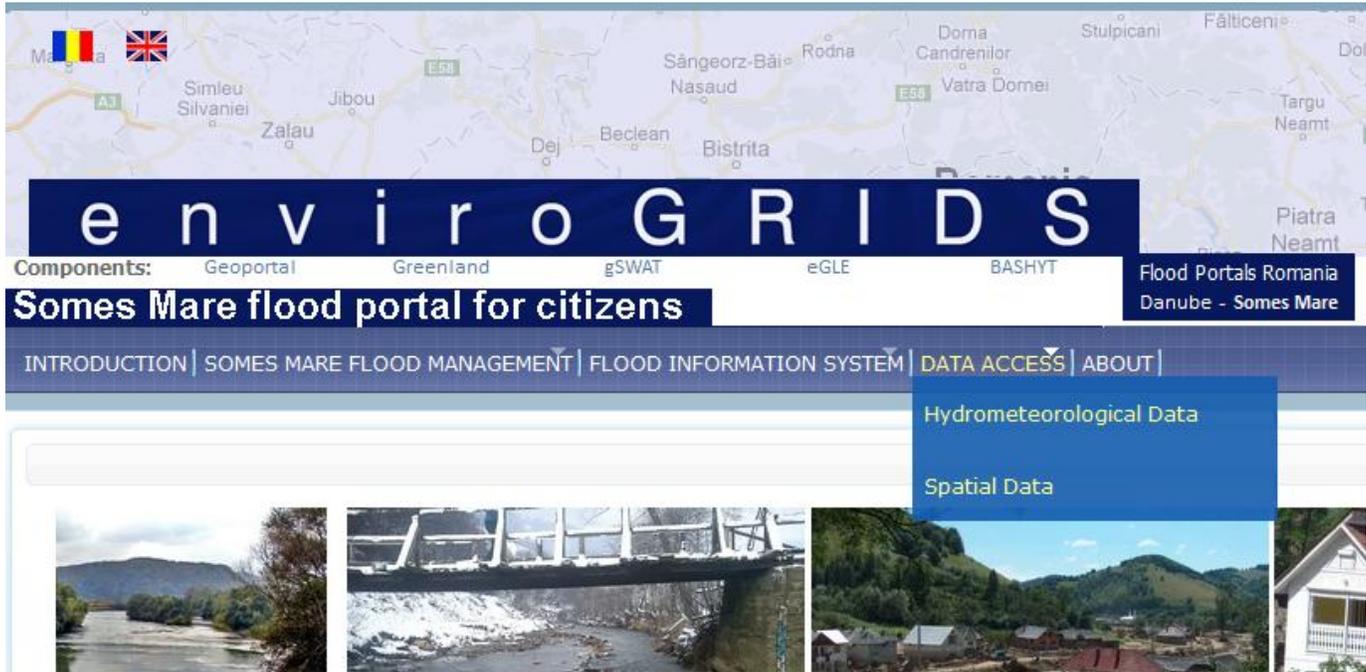

Fig. 5. Somes Mare FIS portal: Data Acess.

The sub section "Hydrometeorological Data" of the data access provides a map based interface for accessing and downloading time series data of precipitation, discharge and temperature for the year 2007 in WaterML 2.0 format. The map based interface (using google maps) presents markers representing monitoring stations that provides the three types of data.   The markers change as user switches from precipitation to discharge or temperature data.

The selection of stations is made possible by clicking on a station marker, using a drop-down list, or by searching for a station nearest to a given address. Once the type of data of the station has been selected, this selection is then displayed in a list box displayed to the right side of the map. Users then need to select in the list box a button to download or view the data that is in WaterML 2.0 format.

Presented in Fig.6 is an example where 4 data items have been selected (two precipitations, one discharge, and one temperature). When "Download" is selected a file type in xml is made available for download. The pre-set file name contains the type of data and name of station (e.g.

Discharge_CHIRALES.xml). When "View" is selected the data are presented in a separate web browser window, as shown in Fig. 7.

With this implementation the available precipitation, discharge and temperature data from the monitoring system of Somes Mare are accessible as web services in WaterML 2.0 format. Since this implementation is for demonstration and testing purposes, such data are provided only for the year 2007. Nevertheless, the implementation allows for other applications to access these web services and test the usage of the data for other purposes.
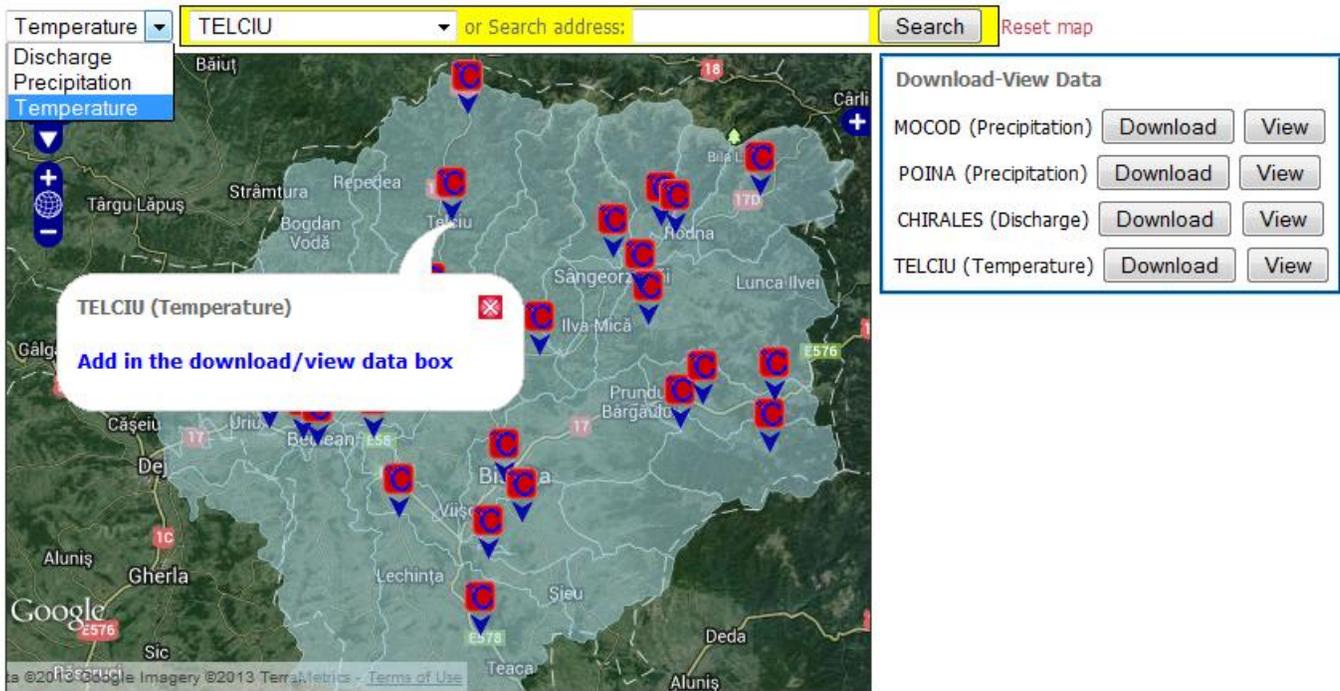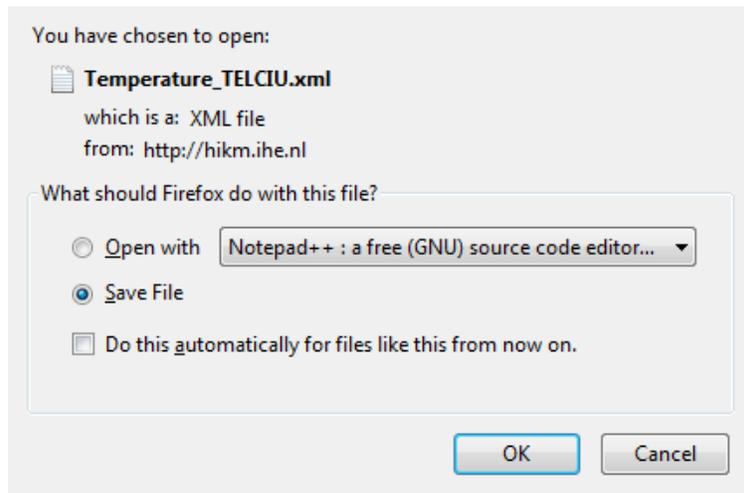
Fig. 6.   Data selection

Fig. 7.   Data web browser viewing and download.

The following is an example of accessing time series data using URL with WFS request.

http://sditest.unesco-ihe.org:8080/GeoserverWaterML/wfs?service=WFS&version=1.1.0&request=getFeature&typeName=om:OM_Observation&outputFormat=gml32&featureID=sbasin.hydrometeo.6.observation.2

V. Conclusions

The WaterML 2.0-GeoServer method has been successfully tested with data from the Somes Mare case study. A web-based front-end interface was developed for FIS to access these time series data. It needs to be noted that the implementation of web services using GeoServer technologies for publishing time series data in WaterML 2.0 format was not straight forward. It required several back-end configurations for achieving the intended functionalities. In general this method for publishing time series data over the web (in WaterML 2.0 format) is practical and promising. However for this method to become more usable for other case study applications there is a need for customized GeoServer package that will set-up a WaterML 2.0 web service (accompanied with a database structure) without any back-end configuration.

Acknowledgment

References

[1] J. Horsburgh, D. Tarboton, M. Piasecki, D. Maidment, I. Zaslavsky, D. Valentine, and T. Whitenack, "An integrated system for publishing environmental observations data," Journal of Environmental Modelling & Software, vol. 24, pp. 879-888, January 2009.

[2] J. A. Sheth and J.Larson, "Federated database systems for managing distributed, heterogeneous, and autonomous databases," Journal of ACM Computing Surveys, vol. 22, pp. 183-236, September 1990.

[3] S. Hickey, and G. Mohan, Participation: from tyranny to transformation? Exploring new approaches to participation in development., Zed Books, 2004.

[4] M.B. Abbott, and A. Jonoski, "The democratisation of decision-making process in the water sector II," Journal of Hydroinformatics,vol. 3(1), pp. 35-47, 2001.

[5] M.S. Reed, "Stakeholder participation for environmental management: A literature review," Journal of Biological Conservation, vol. 141, pp. 2417-2431, October 2008.

[6] R. Ramachandran, S.A. Christopher, S. Movva, X. Li, H.T. Conover, K.R. Keiser, S.J. Graves, R.T. McNider, "Earth Science Markup Language: a solution to address data format heterogeneity problems in atmospheric sciences," Journal of American Meteorological Society, vol. 86 (6), 791–794, June 2005.

[7] S. Cox, Observations and Measurements, Open Geospatial Consortium Best Practices Document, 2006.

[8] I. Zaslavsky, D. Valentine, T.Whiteaker, CUASHI WaterML, Open Geospatial Consortium Discussion paper 07-041r1, 2007.

[9] P. Taylor, OGC WaterML 2.0: Part 1- Timeseries, Open Geospatial Consortium Implemented Standard 10-126r3, 2012.

[10] CSIRO, Publish BOM Observational Data (online), https://www.seegrid.csiro.au/wiki/ASRDC/PublishBomObservationalData, 2012, last accessed July 5, 2013.

[11] A. Almoradie, A. Jonoski, F. Stoica, D. Solomatine, and I. Popescu, " Web-based Flood Information System-: Case study of Somes Mare, Romania," Environmental Engineering and Management Journal, 2013, in press.

[12] A. Jonoski, I. Popescu, A. Almoradie, F. Stoica, S. Teodor, I. Jelev, and D. Gorgan, Functional prototypes of BSC-OS Flood Portals for citizens: Implementation and Evaluation, Final report enviroGRIDS Deliverable D6.11, 2013.

[13] EML Project, Ecological Metadata Language (EML) (online), http://knb.ecoinformatics.org/software/eml/, 2008, last accessed July 5, 2013.