Unifying Hydrological Time Series Data for a Global Water Portal

Pathai Suwannawiwat
Chayakorn Netramai

The Sirindhorn International
Thai-German Graduate School of Engineering
King Mongkut's University of Technology
North Bangkok
Email: pathai@tggs-bangkok.org,
chayakorn.n.sse@tggs-bangkok.org

Michael Utech
Heinz-Josef Schlebusch

KISTERS AG
Email: michael.utech@kisters.de
Schlebusch@kisters.de

Horst Lichter

Research Group Software Construction
RWTH Aachen University
Email: lichter@swc.rwth-aachen.de

***Abstract*—The hydrological observation data is important for hydraulic engineers. They have to use the obtained information to control quality or quantity of water in the river system. Nowadays plenty of hydrological information systems are available. However there are many difficulties to do the parallel accesses between those services, for example, they not only require different approaches to access information, but also return different formats of information as output. Therefore, this work proposes a new design of Data Exchange Template, namely Data Consumer Model (DCM) as a filtering and unify format of the observed time series. This work analyzes on three hydrological information services that have many different factors to proof that proposed solution can work. Finally, the experiment results demonstrate that the DCM is able to access experiment services by simple method. It provides simple and unified format of attributes that used in communication between hydrological systems.**

Keywords*—Web Service, Hydrological Observation, Service Interface, Data Exchange Template

I. Introduction

In order to perform important analyses it has been a long desire to access and retrieve hydrological observation data from different sources by using a simple method. Hydrological observation services have their own characteristics. They do not only require a different approach to access the information but also deliver different format of data to end users. These problems lead to the objective of this paper, which is to find a solution to solve difficulties of accessing hydrological observation services so that users can have a simple way to retrieve data.

This paper proposes the solution of unification problems between hydrological observation services and users. In the context of this research, the unification problem is related to different components of hydrological services such as different response data structures, request filters or service protocols. These issues are important factors that have direct impact to complexity of the service usage. Therefore introducing an intermediate layer is one possible solution which can be used to eliminate unification problems.

An intermediate layer is a simple concept of a medium communication channel between two or more contacts. It can be developed e.g. as a data exchange model which serves a medium for information exchange. The data Exchange Template (DET) supports data exchange by using a standardized format that defines the type of information required in exchange processes. This paper applies the DET concept to develop the Data Consumer Model (DCM) which is the proposed solution. It is developed to be used as an intermediate layer between users and hydrological observation services.

II. Literature Review

DET concept has been used in the development of an exchange template in many fields. There are many papers that apply DET concept in their works: for instance, Object Exchange Model Template (OMET) for GRIDS Distributed Supply Chain Simulation [1], Clinical Document Architecture (CDA) tool, CDA is ANSI-certified standard from Health Level Seven International (HL7.org) [2] and a key exchange and authentication framework along with KEAML Protocol [3]. Therefore, these researches have demonstrated that DET is a useful concept, which is able to be used in different areas like logistics, medical or communication network.

However, there is only one paper [4] that introduced the hydrological observation data template for clients. This model claims to collect observation data from many sources, then converts the received data to the same format as defined in the designed template and then stores the data into the database. This template is represented as a relational data model.

DCM was also developed and being used as a template for hydrological data. [4] and DCM is based on the similar concept of DET. There are many examples of template that are based on the DET concept and this paper will introduce two examples of DET. For instance, Exchange Network [5] provides data templates that are applied from DET for environment information. United States Environmental

12th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), Hua Hin, Thailand, June 24-27, doi: 10.1109/ECTICon.2015.7206974

Protection Agency (EPA) [6] is an organization which provides examples of using data exchange templates[5].

However, DCM in this work has different features compared to the models in [4]. While [4] focuses to the solution of data storage, DCM aims to reduce the hydrological service usages complexity. It means that DCM provides a solution to the difficulties of query process which consists of complexity of services access as well as the complexity of various response formats of data. So we can conclude that a distinctive characteristic of DCM is the feature to provide a simple approach of parallel accesses of services.

III. System Overview

In this work, DCM is used as a data template to reduce the complexity of query processes of hydrological services. It was developed based on the concept of an intermediate layer for both the request and the response side of services. There are two categories of objects within this model. The first part of the model is the query filter that can be used as parameters searching for any data source. The second part is the template for observation data that is being obtained from hydrological services. In general to make use of DCM, developers have to develop applications which embed this model as an internal module, then uses it as data interfaces..

Therefore, after DCM is implemented and used, not only available query filters from request side of DCM will be converted to compatible filters for operations of many hydrological services but also obtained observation data from those services will be converted to the same format as shown in DCM conversion flow in Figure 1. Finally, clients will obtain unified format of observation data as output. The unified format contains less data attributes than original format.

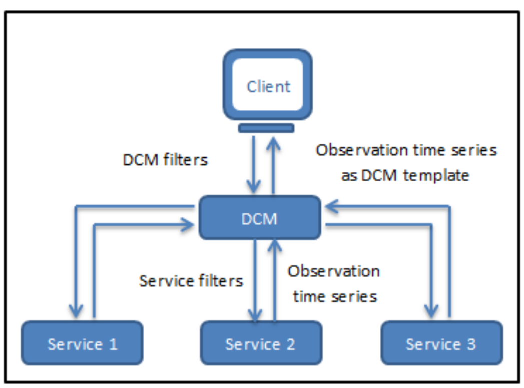

Fig. 1. DCM conversion flow

IV. Implementation

A. Analysis and Selection of Services

The analysis of hydrological observation services was made at the beginning of DCM design phase. It has been found that there are too many numbers of services available. In this work we chose only three services based on the following factors: the differences between *non-public* and *public* standard services, different technologies of services, and different data formats. Therefore, we chose the following services as our experimental samples: Sensor Observation Service (SOS), The United States Geological Survey (USGS) and Kisters Query Service (KIQS).

1) SOS [7]: is a public standardized web service interface for hydrological observation data exchange that is managed by Open Geospatial Consortium (OGC) [8]. It is a component of the Sensor Web Enablement (SWE) [9] standards which is a framework suite for web service interface and communication protocol for sensor network. Its features include managing and retrieving of observation metadata from sensor systems. It also defines a standard for registering new sensors or removing existing sensors, which is a distinctive point of SOS. SOS provides SOAP and HTTP GET with Key Value Pair (KVP) encoding as binding approach. Therefore, SOAP request message is required in query process.

Finally after successfully querying data from SOS web interface by using SOAP message, user will retrieve output data in the Water Markup Language (WML)[10] format. WML is an encoding standard for representing hydrological observation data from SOS. It bases on a combination of Geography Markup Language (GML)[11] and OGC Observations and Measurements standard (O&M)[12] that is represented as XML language.

2) USGS [13]: is a science organization who researches on ecosystems and environment. USGS Water information system has two sources of data provider: *the USGS Water Data Site* and *the USGS Water Service Site*. The USGS Water Data Site delivers a tab-delimited file which has a better performance in terms of data extraction compared to XML and similar formats. It also contains complete observation data. The USGS Water Service Site which is the second source offers more modern and friendly data formats such as SOAP message or REST. However, it contains less observation data compare to USGS Water Data Site. So this paper will focus only on the data provided by the USGS Water Data Site only.

The covered area of USGS is in North America. There are approximately 1.5 million observation stations connected to the USGS water data service in 50 states. However, some stations also investigate and offer different observation data types. USGS does not only collect data relating to quantity, quality, distribution, movement of surface and underground waters but also provides variety time series type such as daily or real time data. Therefore it has become the largest and complex data source based on a nonpublic standard. Despite this complexity, it is still easier to access data from USGS than from SOS.

3) KIQS [14]: is one part of the Water Information System KISTERS (WISKI). WISKI provides features for measuring networks, data capture and evaluation function in every hydrological field. Because KIQS is built based on REST architecture like USGS, all commands from KIQS service will also be available as HTTP GET and POST method, including KVP as binding approach. So that users are allowed to access

12th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), Hua Hin, Thailand, June 24-27, doi: 10.1109/ECTICon.2015.7206974

KIQS service by a URL and obtain observation data as a HTML table or JSON object.

B. Implementation Steps

After the experiment services have been chosen, we researched through their working processes and disparities between them. Our research showed that all services use the concept of observation time series and the station information, which are both related to the collection of the observation data begins at the hydrological stations where raw time series data is being first harvested from the environment and then transferred to hydrological service. After that each hydrological information system will convert raw data into readable format and store it into the database.

We can concluded that stations are the starting point of the query processes. Therefore, station ID can be used as reference for any observation data. When users request for data, the station ID or the interested area is always required in the first step. The request process of observation time series can be summarized as follow:

- Client identifies interested area or station ID.
- Client accesses target service to request for station information or observation time series.
- Target service receives parameters from the client.
- Target service queries data based on received parameters.
- Target service performs response of observation time series or station info then sends back to client.
- Client receive response data.

This request process is illustrated in Figure 2.

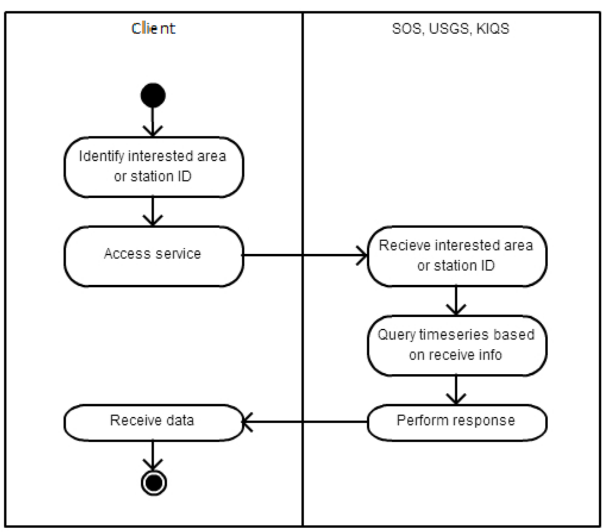

Fig. 2. The common data request processes of hydrological observation services

From the analysis of experiment services, we found that there are two basic commands that exist in all three chosen services: *Get station* command and *Get observation* commands. The *Get station* command is the first requires operation. It always require a station id, a station name or an interested area in the query process then returns the station information back to the clients. The *Get observation* command is the second required operation. It required observation data type or a data period as query filters then it always returns time series of given of observation data back to the clients. Hence we applied the concept of DET with filters from both commands to design request filters of DCM as shown in the working process in Figure 3.

Knowing that any observation data are represented as a time series format, we applied the concept of DET with time series components to design a data template for DCM. This paper focuses on three attributes of time series. These three attributes are able to describe the behavior of time series.

- Time series parameters: used to present observation data type.
- Time series name: providing a time interval of data couple with observation data type such as 15 minutes frequency, daily or monthly of water stream flow.
- Time series ID: identifies a certain time series data, for example, daily time series of water level.

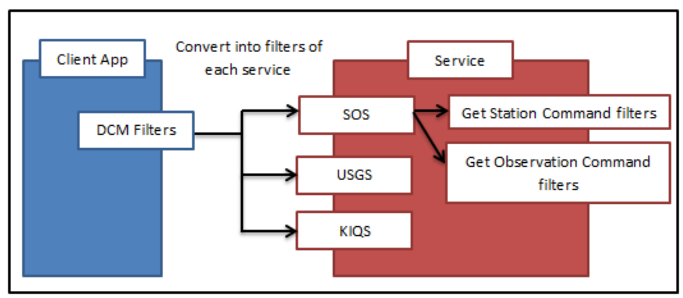

Fig. 3. Working processes of DCM query filters

Figure 4, shows the processes of DCM data template that after conversion of the observation data, we will retrieve "Time series parameters", "Time series name and "Time series ID" for a new template for the output.

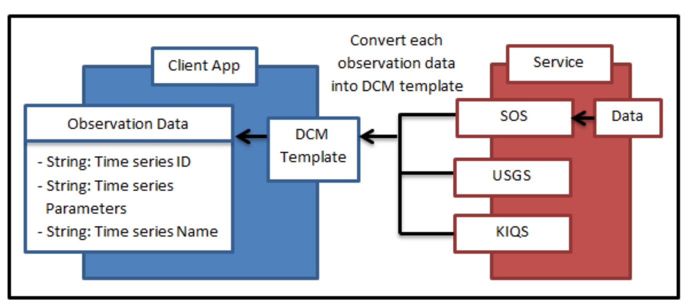

Fig. 4. Working processes of DCM observation data template

C. DCM Components

The DCM filters template consists of filters for both *Get station* command and *Get observation* command as shown in Figure 5. *Get station* command always requires basic parameters like station ID and station area or bounding box, while *Get observation* command requires key attributes of time series like time series ID and time series parameters.

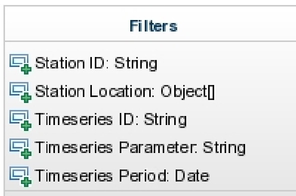

Fig. 5. DCM query filters template

Figure 6 shows the DCM observation data template. Each attribute of station and time series objects is compatible and shared across the experimental services. To make use of the DCM template, clients are allowed to convert this template to array lists where the received observation data is stored in those arrays.

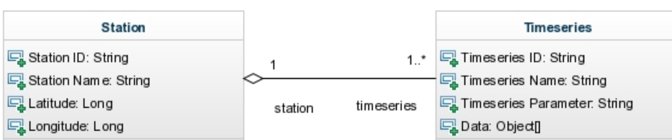

Fig. 6. DCM observation data template

V. Experimental Results

A Java application have been developed to make use of the proposed solution as the data interface to communicate between the experimental services. The *Get station* command and *Get observation* command are implemented in the Java application for experimental purposes.

The experiment uses the developed Java application to measure the data reduction ratio (in percentage) between the input filters and the output data attributes before and after using the proposed DCM template. For example, *agency_cd* attribute from daily value command (DV) from USGS service does not exist in DCM data template. It means the complexity has been reduced already because users do not need to concern about this attribute anymore when they use DCM data template. So, we are able to measure how much complexity of service usage is being reduced (more is better). The experimental sample in Table 1 shows commands from experiment services that have been used. The experimental results in Table 2 and Table 3 show the reduction of actual filters and output data attributes that using the original services compared to the Java application which used the DCM template.

Table 1. A list of experimental commands from each service.

Service	Command	Number of request
SOS	GetFeatureOfInterest	2
	GetDataAvailability (custom operation)	2
	GetObservation	2
USGS	Site information (SI)	2
	Daily value command (DV)	2
	Unit of measurement	2
KIQS	getStationList	2
	getTimeseriesList	2
	getTimeseriesValue	2

Table 2. Reduction result of query filters.

Service	Command	Actual filters (approximate)	Reduction (%)
SOS	Get Station info	4	50%
	Get Observation info	7	29%
USGS	Get Station info	11	82%
	Get Observation info	6	17%
KIQS	Get Station info	13	85%
	Get Observation info	7	29%

Table 3. Reduction result of Observation data attributes.

Service	Command	Actual attributes (approximate)	Reduction (%)
SOS	Get Station info	7	43%
	Get Observation info	21	81%
USGS	Get Station info	7	43%
	Get Observation info	5	20%
KIQS	Get Station info	5	20%
	Get Observation info	7	43%

From the experimental results, we conclude that the proposed DCM is compatible with all three experimental services. Furthermore, DCM is able to communicate with SOS, USGS and KIQS successfully while technical issues like different service protocols have been solved by the Java application. However, a semantic problem was discovered during the experiments. This problem is related to different semantics of the data. The semantic problem is caused by the used vocab and the different language style and it cannot solved by DCM. It requires solutions using AI or statistic methods to deal with it which is outside the scope of this work.

12th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), Hua Hin, Thailand, June 24-27, doi: 10.1109/ECTICon.2015.7206974

VI. CONCLUSION

From the obtained experimental results it can be shown that the proposed solution is able to reduce the hydrological services usage complexity which was the objective of our research. As shown in Table 1 and Table 2 the reduction proportion of both filters and output attributes from original services have been decrease up to 80%. Furthermore, users are able to use the same filters and retrieve the same template of hydrological observation data, which means that DCM offer an easy method to do parallel accesses of hydrological systems to users.

Hence, we can conclude that DCM reduces complexity of services usage by providing a simple and unified format of attributes that used in communication between hydrological systems. Another distinctive feature of this work is an ability to access both information providers that using standardized and non-standardized formats like SOS and USGS.

REFERENCES

[1] Tan, G. ; Wee Ngee Ng ; Taylor, S.J.R. Simulation Symposium, 2002. Proceedings. 35th Annual on April 2002.

[2] Che-Ju Lin ; Der-Ming Liou; Communication Systems, Networks and Applications (ICCSNA), 2010 Second International Conference on June 2010.

[3] Xuebing Qing ; Adams, C. Electrical and Computer Engineering, 2006. CCECE '06. Canadian Conference on May 2006.

[4] R. Maidment, A Data Model for Hydrologic Observations, Paper prepared for presentation at the; University of Texas at Austin, March 7, 2005

[5] Exchange Network, Samples of water Ambient Water Quality, Geospatial Information and Drinking Water Quality data exchange templates. (available at *http://www.exchangenetwork.net/communities-of-interest/water/*)

[6] US Environmental Protection Agency, Monitoring and Assessing Water Quality. (available at *http://water.epa.gov/type/watersheds/monitoring/index.cfm*)

[7] Sensor Observation Service, Schemas and overview (available at *http://www.opengeospatial.org/standards/sos*)

[8] Open Geospatial Consortium, OGC standards list (available at *http://www.opengeospatial.org/standards/is*)

[9] Sensor Web Enablement, list of related standard and overview (available at *http://www.opengeospatial.org/ogc/markets-technologies/swe*)

[10] Water Markup Language, WaterML official schemas and overview (available at *http://www.opengeospatial.org/standards/waterml*)

[11] Geography Markup Language, GML official schemas and overview (available at *http://www.opengeospatial.org/standards/gml*)

[12] Observations and Measurements, O&M official schemas and overview (available at *http://www.opengeospatial.org/projects/groups/om2.0swg*)

[13] The United States Geological Survey, Web interface of Water Data Site service (available at *http://waterdata.usgs.gov/nwis/rt*)

[14] Kisters Query Service, List of wiski modules, Web interface of KIQS service (available at *http://www.kisters.net/wiski-modules.html*)

12th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), Hua Hin, Thailand, June 24-27, doi: 10.1109/ECTICon.2015.7206974