CUAHSI

universities allied for water research

CUAHSI HIS DATA CART SPECIFICATION

Version 1.0

June 7, 2010

by:

Tim Whiteaker
Center for Research in Water Resources
The University of Texas at Austin

Dean Djokic
Environmental Systems Research Institute

Distribution

Funding

Funding for this document was provided by the Consortium of Universities for the Advancement of Hydrologic Science, Inc. (CUAHSI) under National Science Foundation Grant No. EAR-0622374. In addition, much input and feedback has been received from the CUAHSI Hydrologic Information System development team. Their contribution is acknowledged here.

Table of Contents

Version History

April, 2009 – "Thematic Dataset Tables" designed by HIS team
July, 2009 – Dart Cart design generated from Thematic Dataset Tables by UT-Austin/ESRI
March, 2010 – Data Cart design refined by HIS team and ESRI
May 19, 2010 – After discussion on the HIS call, Location and Variable were added back in, along with a field that indicates the version of the WaterOneFlow service. We currently need Location and Variable to handle complex situations like NWIS, which has several "overloads" for those two fields.
June 2, 2010 – Location and Variable were made "optional" fields. While these fields may be provided by the cart originator, clients are expected to be able to reassemble these fields by reading data from other fields in the cart. This will be necessary if the user wants to tweak the GetValues request for a time series, or if the syntax for WaterOneFlow web service method requests change with the release of subsequent versions of WaterOneFlow or the WaterML specification.

Data Cart Design

A data cart is a structure that provides information needed for a client to access each time series described in the cart, as well as a few important fields for helping the human to get sense of what the time series in the cart represent and to assist in selecting a subset of the time series for retrieval. Each entry in the data cart identifies a single time series that can be retrieved via a GetValues request to a WaterOneFlow service.

The data cart could be implemented in a number of formats. For the purposes of this discussion, assume the data cart is implemented as a table. The fields in the data cart are described in Table 1. The field names are restricted to 10 characters or less for shapefile support. For text fields, the maximum number of characters allowed for the field is also provided.

Required fields are written in **bold text.** The remaining fields are there to help you assess the time series before actually requesting the data. For example, you might only want to download time series for observations recorded using a specific measurement method. In order to get a better sense of the data, you may also one to perform some quick summaries on the fields in the data cart to identify all of the unique methods used or to determine all of the unique site locations (handy when a site measures more than one variable and thus appears in the table more than once).

Table 1 Data Cart Fields

Field Name (Field Type)	Definition	Example
ServCode (Text - 50)	Network prefix for site codes used by the WaterOneFlow service, giving the context within which the site code applies	CCBay
SiteCode (Text - 50)	Unique text identifier for a site within a given WaterOneFlow service	H1
SiteName (Text - 255)	Name of a site	Hypoxia_1

VarCode (Text - 50)	Unique text identifier for a variable within a given WaterOneFlow service	DOC
VarName (Text - 255)	Name of a variable	Dissolved Oxygen Concentration
VarUnits (Text - 50)	Units of measure for the variable	milligrams per liter
Vocabulary (Text - 50)	Vocabulary prefix for variable codes giving the context within which the code applies	CCBay
Ontology (Text – 50)	Unique name for the ontology containing the concept to which the given variable has been mapped	CUAHSI Variable Ontology v1.26
Concept (Text - 50)	Leaf concept keyword from the ontology to which this variable applies	dissolvedOxygen
ValueCount (LongInt)	Number of time series values for the variable at the site for the given time period	270
StartDate (Date)	Start date and time for the time period of the variable at the site	5/3/94 8:40 AM
EndDate (Date)	End date and time for the time period of the variable at the site	8/31/06 11:26 AM
Latitude (Double)	Latitude of the site location in decimal degrees (WGS_1984); for polygons can be *NULL*	27.814
Longitude (Double)	Longitude of the site location in decimal degrees (WGS_1984); for polygons can be *NULL*	-97.141
IsRegular (ShortInt)	1 (TRUE) if variable is measured/calculated regularly in time; 0 (FALSE) otherwise	0
TimeUnits (Text - 50)	For regular data, the time step and time units give the length of time between measurements, e.g., 1 day, 6.5 hrs, 1 month	Day
TimeStep (Double)	For regular data, the time step and time units give the length of time between measurements, e.g., 1 day, 6.5 hrs, 1 month	1
DataType (Text - 50)	Type of data	Value, Average, Maximum, Minimum, StandardDeviation
Medium (Text - 50)	Medium in which the variable applies	Surface Water
MethodID (Integer)	Unique ID within a WaterOneFlow service for the method used to measure the variable	1
Method (Text - 255)	Description of the method used to measure the variable	Multiprobe measurement
QCLevelID (Integer)	Unique ID within a WaterOneFlow service for the quality control level of the time series	0
QCLevel (Text - 50)	Description of the quality control level of the time series	Raw Data
SourceID (Integer)	Unique ID within a WaterOneFlow service for the original source of the data	1
SourceName	Name of the original source of the data	Texas A&M University Corpus Christi

(Text - 255)		
LocType (Text – 25)	Type of service – indicates how the Location parameter of a WaterOneFlow.GetValues call should be formatted	SiteCode LatLongBox LatLongPoint
ServType (Text – 10)	Type of endpoint, REST, SOAP	SOAP
XLL (Double)	For point data, Longitude of the point. For data defined by a lat/lon box, western longitude of the box	-97.141
YLL (Double)	For point data, Latitude of the point. For data defined by a lat/lon box, southern latitude of the box	27.814
XUR (Double)	For data defined by a lat/lon box, eastern longitude of the box; otherwise can be *NULL*	-93.5
YUR (Double)	For data defined by a lat/lon box, northern latitude of the box; otherwise can be *NULL*	30.2
Location (Text - 255)	Properly formatted location parameter to pass to WaterOneFlow.GetValues	CCBay:Hypoxia_1 GEOM:BOX(-97.141 27.814,-93.5 30.2) GEOM:POINT(-97.141 27.814)
Variable (Text - 255)	Properly formatted variable parameter to pass to WaterOneFlow.GetValues	CCBay:DOC NWISDV:00060/DataType=Maximum
ReqsAuth (ShortInt)	Request authorization. 1 (TRUE) if authorization for download is required; 0 (FALSE) otherwise	0
WaterMLURI (Text - 255)	URI of WaterOneFlow service WSDL	http://data.com/WoF/ /cuahsi_1_0.asmx?WSDL
WofVersion (Text - 15)	Version of the WaterOneFlow service	1.0
WFSURI (Text - 255)	URI of web feature service showing site locations	http://data.com/WFSServer
WMSURI (Text - 255)	URI of web mapping service related to the data	http://data.com/WMSServer
DAccessURI (Text - 255)	URI of Data Access Service, which provides REST querying capabilities for WaterOneFlow, user management, data cart management, and more	http://data.com/DataService

Feature Classes To Represent Geometry

It may be useful to provide a separate feature class of just the site locations in the same package as the data cart. There are two feature class representations – point and polygon (box) to reflect the two geometric options for definition of the "observational" feature. The fields in these feature classes are shown in Table 2 and Table 3.

Table 2 Point Feature Class Fields

Field Name (Field Type)	Definition	Example

HydroID (Integer)	Unique identifier for the feature in the geodatabase	101
ServCode (Text - 50)	Network prefix for site codes used by the WaterOneFlow service, giving the context within which the site code applies	CCBay
SiteCode (Text - 50)	Unique text identifier for a site within a given WaterOneFlow service	H1
SiteName (Text - 255)	Name of a site	Hypoxia_1
Latitude (Decimal)	Latitude of the site location in decimal degrees	27.814
Longitude (Decimal)	Longitude of the site location in decimal degrees	-97.141

Table 3 Polygon (box) Feature Class Fields

Field Name (Field Type)	Definition	Example
HydroID (Integer)	Unique identifier for the feature in the geodatabase	201
ServCode (Text - 50)	Network prefix for site codes used by the WaterOneFlow service, giving the context within which the site code applies	CCBay
SiteCode (Text - 50)	Indicates geometry of the feature	BOX(-111 41,-104 45)
SiteName (Text - 255)	Name of a polygon area	Model Grid Cell 201
XLL (Double)	For point data, Longitude of the point. For data defined by a lat/lon box, western longitude of the box	-111.0
YLL (Double)	For point data, Latitude of the point. For data defined by a lat/lon box, southern latitude of the box	41.0
XUR (Double)	For data defined by a lat/lon box, eastern longitude of the box	-104.0
YUR (Double)	For data defined by a lat/lon box, northern latitude of the box	45.0

When including these additional feature classes for point and polygon locations, a FeatureID field can be added to the DataCart table, which points to the HydroID field in the point and polygon feature classes. This is used to establish a one-to-many relationship between the point feature class (one) and data cart (many), or the polygon feature class (one) and data cart (many) (Figure 1).

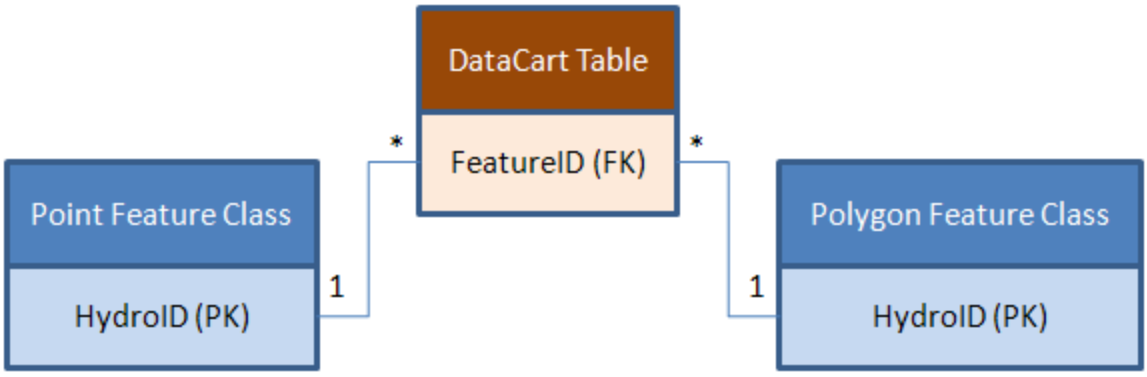

Figure 1 Data Cart Relationships

Another possible implementation is for the data cart itself to be a feature class, in which case there may be spatially coincident features if a given feature has more than one time series in the data cart. Another possible limitation is when both point and polygon features are defined in the cart. Standard GIS data types like shapefiles do not allow multiple geometry types to be stored in a single feature class.

Additional feature classes, tables, and files could be packaged with the data cart to aid the user in interpreting the data, such as political boundaries, layer files, and even ArcMap documents with predetermined layouts.

Data Cart Metadata

A data cart should include metadata about the cart itself, such as who created the cart and for what purpose. This will be necessary if users are to share carts and register them with CUAHSI Online. Data cart metadata has not yet been formulated by the HIS team.

Client Expectations

Client applications read items in a data cart and download the time series described by those items. Client applications are expected to be able to

- Make web service requests
- Determine the version of a WaterOneFlow web service and make data requests accordingly
- Assemble items from the required fields of the data to generate the parameters required by the WaterOneFlow web service.

Note
Client applications may opt to read WaterOneFlow parameters directly from the Location and Variable fields in the data cart. However, this practice should be avoided because the data cart originator may not have populated these fields, and the values in these fields may result in a failed web method request if the WaterOneFlow web service has been updated to a new version since the data cart was created.

EXAMPLE – PARADISE FLOW

The following example shows a data cart in tabular format, created for streamflow in the Little Bear River near the city of Paradise, Utah. Streamflow time series for three sites are identified from three different WaterOneFlow web services (and in this case, data providers).

Because the data cart has so many fields, the cart will be displayed in multiple tables in this document. The ServCode field will be repeated in each table to differentiate each data row.

Note
Some of the "required" fields are missing below. That is because the table below shows real data and these fields are not yet returned by the WaterOneFlow services providing access to the data. This should be remedied!

ServCode	SiteCode	SiteName
LittleBearRiver	USU-LBR-Confluence	Little Bear River below Confluence of South and East Forks near Avon, Utah
EPA	UTAHDWQ:4905750	E FK LITTLE BEAR R AB CNFL / S FK LITTLE BEAR R
NWISDV	10105900	LITTLE BEAR RIVER AT PARADISE, UT

ServCode	VarCode	VarName	VarUnits	Vocabulary	Ontology	Concept
LittleBearRiver	USU44	Discharge	cubic feet per second	LBR	CUAHSI 1.0	Streamflow
EPA	239-1	Flow	cfs	EPA	CUAHSI 1.0	Streamflow
NWISDV	00060	Discharge, cubic feet per second	cubic feet per second	NWISDV	CUAHSI 1.0	Streamflow

ServCode	ValueCount	StartDate	EndDate	Latitude	Longitude
LittleBearRiver	10594	11/14/2007 5:30 PM	6/22/2008 11:30 AM	41.536088	-111.830455
EPA	92	5/8/1990 4:40 PM	5/17/2006 3:59 PM	41.52972031	-111.8122253
NWISDV	6404	10/1/1991	6/3/2010	41.57548904	-111.855217

ServCode	IsRegular	TimeUnits	TimeStep	DataType	Medium
LittleBearRiver	1	minute	30	Average	Surface Water
EPA	0			Instantaneous	Surface Water
NWISDV	1	day	1	Average	Surface Water

ServCode	MethodID	Method
LittleBearRiver	29	Discharge derived from water level measurements using a site specific stage-discharge relationship.
EPA		UTAHDWQ:FIELD MEASURES
NWISDV		

ServCode	QCLevelID	QCLevel	SourceID	SourceName
LittleBearRiver	2	Derived products	2	Utah State University Utah Water Research Laboratory
EPA				Utah Division of Water Quality
NWISDV		Unknown		USGS

ServCode	LocType	ServType	XLL	YLL	XUR	YUR
LittleBearRiver	SiteCode	SOAP	-111.830455	41.536088		
EPA	SiteCode	SOAP	-111.8122253	41.52972031		
NWISDV	SiteCode	SOAP	-111.855217	41.57548904		

ServCode	Location	Variable	ReqsAuth	WaterMLURI
LittleBearRiver	LittleBearRiver:USU-LBR-Confluence	LBR:USU44	0	http://icewater.usu.edu/littlebearriver/cuahsi_1_0.asmx?WSDL
EPA	EPA:UTAHDWQ:4905750	EPA:239-1	0	http://river.sdsc.edu/wateroneflow/EPA/cuahsi_1_0.asmx?WSDL
NWISDV	NWISDV:10105900	NWISDV:00060/DataType=Average	0	http://river.sdsc.edu/wateroneflow/NWIS/DailyValues.asmx?WSDL

ServCode	WofVersion	WFSURI	WMSURI	DAccessURI
LittleBearRiver	1.0			
EPA	1.0			
NWISDV	1.0			

The locations of these sites are shown in the crude Google Maps image in Figure 2.

In order to download the time series for USU-LBR-Confluence, the client application first looks at the WaterMLURI field for the row in the data cart with "SiteCode='USU-LBR-Confluence'" to find the URL of the web service WSDL. The value of the WaterMLURI in this case is "http://icewater.usu.edu/littlebearriver/cuahsi_1_0.asmx?WSDL." The WSDL is a document on the Internet that defines the methods available from a given web service. The client interprets the WSDL to figure out the location of the WaterOneFlow service and other information such as the methods available on the service or perhaps the service version (for WaterOneFlow version 1.1 and earlier).

Next, the client makes a GetValues request to ask the service for the time series. In version 1.0, a WaterOneFlow GetValues request requires five parameters from the client: location, variable, startDate, endDate, and authToken. The client assembles these parameters by reading fields in the data cart. The structure of these parameters is described at http://river.sdsc.edu/wiki/CUAHSI WaterOneFlow 1_0 API.ashx and http://river.sdsc.edu/wiki/GenericODWS v1_1 features.ashx.

Table 4 shows how the GetValues parameters are constructed from the fields in the data cart for the USU-LBR-Confluence site. Note that the variable parameter that is constructed by the client application is different than the value in the Variable field present in the data cart. It turns out that for this particular site, only one time series is available for the USU44 variable, and so the additional filters of SampleMedium and DataType are not necessary. However, clients have no way of knowing this, and so they should always construct the full variable parameter.

Figure 2 Location of sites near Paradise, UT, which measure streamflow

Table 4 WaterOneFlow GetValues parameters for the time series at USU-LBR-Confluence

Parameter	Data Cart Fields Used	Result	Comments
location	ServCode, SiteCode, LocType, XLL, YLL, XUR, YUR	LittleBearRiver:USU-LBR-Confluence	Since LocType is "SiteCode" in this case, the ServCode is concatenated with the SiteCode

variable	Vocabulary, VarCode, Medium, DataType, MethodID, SourceID, QCLevelID	LBR:USU44/DataType=Average/SampleMedium=Surface Water/METHODID=29/SOURCEID=2/QUALITYCONTROLLEVELID=2	
startDate	StartDate	2007-11-14T17:30:00	Dates are written in ISO time format
endDate	EndDate	2008-06-22T11:30:00	Dates are written in ISO time format
authToken	ReqsAuth	*[left blank]*	Because the service does not require an authorization token, this field is left blank

CONCLUSIONS

This document has specified a data cart design for describing time series, which gives client applications all of the information that they need to make requests to download those time series. There are some areas for improvement, such as handling real-time or forecast data. However, this represents a step forward and will serve as a point of departure for refining both the design of data carts and also future versions of WaterOneFlow and WaterML.