Appendix: HydroBase Input Type

2006-11-06, Acrobat Distiller

Overview

The State of Colorado's HydroBase database stores a variety of time series data. The time series conventions described here, in particular for time series identifiers, are consistent for major CDSS software components including TSTool, StateView/CWRAT, StateDMI, and StateMod GUI. This allows for consistent features and sharing of data between software tools.

The current database design splits time series into three main categories:

1. Data related to structures or administrative data maintained by the State of Colorado (e.g., diversions, reservoirs). Structure locations are typically identified using a water district identifier (WDID), consisting of a two digit State of Colorado water district number and a trailing structure identifier (which in the past was four digits but has been increased to five or more digits to support longer identifiers). Although a single WDID identifier is used when identifying time series, the separate WD and ID fields are generally needed to find information in HydroBase.
2. Data for stations, consisting mainly of location information and time series (e.g., NOAA precipitation data, USGS streamflow). Station locations are typically identified using a station identifier from the data source. For example, stations can use a USGS identifier, a State of Colorado Satellite Monitoring System abbreviation, or other identifier.
3. Data recorded at locations that are not stations or structures. For example, Water Information Sheet (WIS) are daily spreadsheets used to administer water. Although WIS contain data values for structures an stations, the time series are extracted from database tables that are not directly associated with structure or station database tables. Other examples include Colorado and national agricultural crop statistics.

A structure or station may have more than one identifiers depending on the number of agencies involved with data collection, etc. For example, a reservoir may have a State of Colorado WDID because it has water rights, a Bureau of Reclamation identifier, a US Geological Survey identifier, and a second State of Colorado identifier because real-time data are collected. HydroBase collects data from many sources; however, the State has not attempted in all cases to cross-references the identifiers. For example, a streamflow station may have a partial time series record with a "USGS" data source and identifier and a partial time series record with a "DWR" (Division of Water Resources) data source and identifier – the user must recognize that this may be the same station, under different management at different times.

HydroBase is updated for release to the public approximately once per year, although internal updates may occur year-round. Time series are used with CDSS (Colorado's Decision Support Systems) applications and follow basic time series standards when used by TSTool and other software.

HydroBase and Standard Time Series Properties

The standard time series identifier format for HydroBase time series is of the form:

```
Location.DataSource.DataType.Interval~HydroBase
```

Due to the variety of data types, sources, and formats in HydroBase, time series properties can be set a number of ways. General guidelines are as follows:

- The location part of the time series identifier is set to a station or structure identifier, which is typically the identifier used by the managing agency. For example, USGS stream gages will use the 8-digit USGS identifier and State of Colorado diversions will use a structure WDID.
- The source part of the time series identifier corresponds to the current source of the data. For example, if the current provider for a time series is the USGS, then the data source will be USGS. If the State of Colorado has at some point taken over maintenance of a station from the USGS, then the data source will be DWR. Individual data records may indicate a variety of data sources. The convention in HydroBase is to store the data records under the current data source, rather than force the user to query more than one time series and merge the time series. If, however, a station has moved, then separate time series will typically be stored, likely under different identifiers.
- The data type part of the time series identifier as much as possible uses the "measurement type" information in HydroBase or a readable and reasonable data type phrase. For example "Precip" is a measurement type for station data and "DivTotal" (diversion total) is a measurement type for diversion data. In some cases, especially with real-time data, the data type may not exactly match HydroBase. For example, HydroBase uses a measurement type "RT_Rate" for multiple stream related data types. TSTool uses a data type of "Streamflow". In the past, TSTool and other software used data types that did not as closely match the measurement types in HydroBase. For example, daily streamflow was identified as QME (a National Weather Service notation) because that is how it was defined in CRDSS modeling efforts. The table at the bottom of this appendix describes all available HydroBase data types and provides guidance for upgrading from old data types.
- Data intervals are set based on the tables that are being queried. In most cases, a regular interval like DAY or MONTH is used. IRREGULAR is used for real-time data because there is currently no way to know without doubt what the regular data interval is (e.g., 15MIN). Data that are measured infrequently (e.g., reservoir field measurements) are typically stored as a regular interval time series with interval DAY. This allows more flexibility in data processing and filling.
- In older versions of TSTool, the scenario part of the identifier was sometimes used to supplement the data type information. For example, real-time flow data in the database has a number of attributes (`Streamflow`, `RT_Rate`, `DISCHRG`) that cannot easily fit into the standard time series identifier. The current version of TSTool uses datatype-subdatatype where necessary and generally does not use the scenario for normal time series identifiers (WIS time series are an exception) and this field is being reserved to possibly indicate historical data, filled data, etc.
- Units are set based on the database table definitions.
- Period of record is set based on the available database contents. Periods are typically not determined by checking the data because this would require querying large amounts of data. When listing time series, periods are normally determined from summary information available in the database. In some cases, the period of record information is not saved at a precision sufficient to accurately represent the true period (e.g., the database may indicate data for years but not months). Therefore, the true period will only be available when data are actually queried.
- Missing data are typically set to `-999` in time series but are typically stored as nulls in the database.
- The input type of the time series identifier may not be used for older applications. The new convention is being phased in and uses an input type of `HydroBase` (e.g.,

`12345678.USGS.QME.DAY~HydroBase`). If multiple HydroBase connections are needed, the input name may also be added (e.g., `12345678.USGS.QME.DAY~HydroBase~ServerName`), although this capability is only in the evaluation stage.

- The time scale for data (whether accumulated [`ACCM`], instantaneous [`INST`], or mean [`MEAN`]) is not automatically determined from the data type and interval.

The following tables present a summary of time series identifier fields for the HydroBase data types. Data sources may be added and/or removed with data updates. Data types are listed by major group and are alphabetized by the data type description within the group. The time scale is provided to facilitate data use, in particular when changing the time interval.

HydroBase Time Series Types and Standard Time Series Identifier Fields
Agricultural Crop and Livestock Data

Data Group	Data Type Description	Location	Data Source	Data Type	Available Intervals and Time Scale	Comments
Agricultural/ CASS	Colorado Agricultural Statistics Service crop area harvested	County Name	CASS	CropAreaHarvested-Commodity_Practice Commodity and practice are from available values in HydroBase.	Year INST	See NASS data for orchards, pasture, and vegetables. Perennial crops usually have only harvested value.
	CASS area planted	County Name	CASS	CropAreaPlanted-Commodity_Practice Commodity and practice are from available values in HydroBase.	Year INST	Annual crops should have planted value but use maximum of planted and harvested if necessary.
	CASS livestock head	County Name	CASS	LivestockHead-Commodity_Type Commodity and type are from available values in HydroBase.	Year INST	For each commodity (e.g., sheep), multiple types (e.g. sheep at various maturity levels).
Agricultural/ GIS	CDSS irrigated lands assessment result. See also Diversion Comments below.	WDID	CDSSGIS	CropAreaAllIrrigation-CropType CropAreaDrip-CropType CropAreaFlood-CropType CropAreaFurrow-CropType CropAreaSprinkler-CropType CropType is taken from available values in HydroBase.	Year INST	Data are only available for years where DSS projects or data refreshes have occurred. Partial data for intermediate years may be available in spatial data layer attributes but not HydroBase. Data are available for lands served by surface water structures, listed by crop/year/irrigation type.
Agricultural/ NASS	CropArea	County Name	NASS	CropArea-Commodity Commodity is taken from available values in HydroBase.	Year INST	See CASS data where available. NASS does not distinguish between planted and harvested. NASS data are useful for orchards, pasture, and vegetables, which may not be reported in CASS.

HydroBase Time Series Types and Standard Time Series Identifier Fields (Climate Data)
Climate Group Table 1 of 2

Data Group	Data Type Description	Location	Data Source	Data Type	Available Intervals and Time Scale	Comments
Climate	Evaporation (Pan)	Station ID	NOAA	EvapPan Old (obsolete) data type was EPAN.	Day ACCM, Month ACCM	
	Frost Dates (derived from temperatures)	Station ID	COAGM, NOAA	FrostDateL28S, FrostDateL32S, FrostDateF28F, FrostDateF32F Old (obsolete) data type was FrostDate or FrostDates.	Year INST	Time series in software are the Julian day of the year (1-366) to allow graphing, filling, and manipulation.
	Precipitation	Station ID	COAGM, NOAA	Precip Old (obsolete) data type was PTPX.	Day ACCM, Month ACCM, Irregular ACCM	Irregular data are real-time increments.
	Snow (accumulation on ground during interval).	Station ID	NOAA	Snow Old (obsolete) data type was SNOG.	Day ACCM, Month ACCM	
	Snow course depth and snow water equivalent	Station ID	SCS	SnowCourseDepth, SnowCourseSWE Old (obsolete) data type was SnowCrse, SNWE.	Day INST	Values are recorded on a day, with one or more times a month.
	Solar radiation	Station ID	COAGM	Solar Old (obsolete) data type was RADS.	Day ACCM	
	Temperature (instantaneous)	Station ID	various	Temp	Irregular INST	
	Temperature (maximum)	Station ID	COAGM, NOAA	TempMax Old (obsolete) data type was MaxTemp, TAMN.	Day INST	
	Temperature (mean of maximum daily values)	Station ID	COAGM, NOAA	TempMeanMax Old (obsolete) data type was MaxTemp, TAMX with monthly interval.	Month MEAN	
	Temperature (mean)	Station ID	COAGM, NOAA	TempMean Old (obsolete) data type was MeanTemp, TAVG.	Month MEAN	
	Temperature (minimum)	Station ID	COAGM, NOAA	TempMin Old (obsolete) data type was MinTemp, TAMN.	Day INST	
	Temperature (mean of minimum daily values)	Station ID	COAGM, NOAA	TempMeanMin Old (obsolete) data type was MinTemp, TAMN with monthly interval.	Month MEAN	

HydroBase Time Series Types and Standard Time Series Identifier Fields (Climate Data)
Climate Group Table 2 of 2

Data Group	Data Type Description	Location	Data Source	Data Type	Available Intervals and Time Scale	Comments
Climate	Vapor pressure (mean daily)	StationID	COAGM	VaporPressure Old (obsolete) data type was VP, MVP.	Day MEAN	
	Wind run	Station ID	AGRO, COAGM	Wind Old (obsolete) data type was UDIS.	Day ACCM	

HydroBase Time Series Types and Standard Time Series Identifier Fields (Demographic Data)

Demographic data are related to human population. See the Agricultural Data above for livestock population.

Data Group	Data Type Description	Location	Data Source	Data Type(s)	Available Intervals and Time Scale	Comments
Demographics	Human population (persons)	Area_type-Area_name The type indicates whether a county, municipality, state, etc. The name agrees with the type. The combination defines a unique location.	(blank) This could be assumed from the Pop_type part of the data type; however, the data source is not readily available in HydroBase.	HumanPopulation-Pop_type The population type is Census, Estimated, etc.	Year INST	See CDSS documents for information on how population estimates are determined.

HydroBase Time Series Types and Standard Time Series Identifier Fields (Diversion Data)

Data Group	Data Type Description	Location	Data Source	Data Type	Available Intervals and Time Scale	Comments
Diversion May include records for reservoir and well structures, as per State of Colorado administration practices. See also reservoir data.	Diversion Class (showing water color)	WDID	DWR	DivClass-SFUT Old (obsolete) data type was DQME, Div, or Diversion.	Day MEAN, Month INST or ACCM, Year INST or ACCM	SFUT is encoded as: S:s F:f U:u T:t s = source f = from u = use t = type Annual values are for irrigation year (Nov-Oct).
	Diversion Comment (the acreage for a diversion and string data flag indicating whether a structure irrigated in a year)	WDID	DWR	DivComment	Year INST or ACCM	The numerical time series value is set to the acreage for the year. The data quality flag is set to the HydroBase *diversion_comment.not_used* flag. Therefore, this time series can be used to extract total acreage for a structure and determine if diversions should be zero for a year. Annual values are for irrigation year (Nov-Oct).
	Diversion Total (sum of all DivClass records for a structure).	WDID	DWR	DivTotal Old (obsolete) data type was DQME, Div, or Diversion.	Day MEAN, Month INST or ACCM, Year INST or ACCM	Annual values are for irrigation (Nov-Oct) year.

The above table summarizes how diversion records are available as time series. However, to determine a complete diversion time series, it is necessary to understand the various ways that diversion records can be stored. See also the **State of Colorado's Water Commissioner Manual**.

Raw data observations for a diversion structure are stored as one or more of the following forms in HydroBase:

- Daily water class time series. These data are recorded using irrigation year (November to October). If one or more values have been entered in a month, then HydroBase will include a full month of data. Days at the beginning of the irrigation year that have no observed values at the start of the year should be considered to be zero, regardless of values found in previous irrigation years. Once an observation occurs, then days within the month where an observation was not recorded are set to the last observed value. Therefore, if an irrigation year contains at least one value, that irrigation year will have at least one month of values (with no missing in the month). To preserve space in HydroBase, months with no observations are not included in the daily data in the database. If a year has no observation, then no data are available in HydroBase for the year and a determination of whether the data values should be zero or other must be determined using other data (see below) or engineering judgment. **TSTool and StateView by default implement the carry-forward procedure within irrigation years.**
- Diversion comments. Diversion comments may be included for an irrigation year. The not_used flag indicates if a diversion was not used in a year. If this is the case, then daily diversion records should not be available and a zero value can be assumed for the water year. **TSTool and StateView DO NOT by default use diversion comments when providing daily or monthly time series.**
- Infrequent water class. Infrequent water class values can be entered as an annual value for the irrigation year, or as a monthly value. The data can be accessed as time series in TSTool, although no specific capabilities have been implemented to supplement the daily or monthly time series.

Processed (derived) data records are created as follows:

- Daily total diversion. Daily water class values are accumulated to daily total records. Similar to the daily water class, any month that has at least one value will result in a month with no missing data. To preserve space in HydroBase, only months that include an observation are included in HydroBase. Other months in the same irrigation year should be carried forward. Irrigation years with no observation have no records in HydroBase and a determination of whether the data values should be zero or other must be made using other data (see below) or engineering judgment. **TSTool and StateView by default implement the carry-forward procedure within irrigation years.**
- Monthly water class. Monthly water class is computed by converting the daily water class values (average CFS) to ACFT for each day of the month, and adding the values. Because of the way that daily data are treated, a month will either have all daily values or none. A month with no data will have its value set to missing in the database. Full irrigation years with no observation will result in a full year of missing values, and a determination of whether the data values should be zero or other must be determined using other data (see below) or engineering judgment. Unlike daily data, monthly diversion records are included in HydroBase for the full data period. Full years of missing values may be included in the database.
- Monthly total diversion. This is derived using the same procedure as monthly water class; however, the daily total diversion is used as input.
- Infrequent data are not considered when producing the monthly total time series.

Therefore, to determine a complete time series, the following must be performed, using TSTool or other software:

Daily time series:

1. Read the daily time series from HydroBase. The default in TSTool and StateView is now to carry forward daily diversion time series within the irrigation year.
2. Utilize the diversion comments to set additional years of data to zero. Using diversion comments is an option with TSTool and StateDMI time series read commands.
3. For years with no data, use an appropriate fill technique. If it is known that the ditch did not operate, then zeros should be used. If it is known that the ditch did operate, use historical averages or some other method to fill the data.
4. HydroBase infrequent diversions could be used to supplement the data, but currently there is no software to help users with this process.

Monthly time series:

1. Read the monthly time series from HydroBase. Any irrigation year with at least one daily observation results in 12 monthly time series values.
2. Utilize the diversion comments to set additional years of data to zero. Using diversion comments is an option with TSTool and StateDMI time series read commands.
3. For years with no data, use an appropriate fill technique. If it is known that the ditch did not operate, then zeros should be used. If it is known that the ditch did operate, use historical averages or some other method to fill the data.
4. HydroBase infrequent diversions could be used to supplement the data, but currently there is no software to help users with this process.

Yearly time series:

1. Infrequent time series can be read by TSTool and can supplement the above data. However, currently there is no software to help users with this process. General TSTool commands must be used as appropriate.

HydroBase Time Series Types and Standard Time Series Identifier Fields (Hardware Data)

Data Group	Data Type Description	Location	Data Source	Data Type(s)	Available Intervals and Time Scale	Comments
Hardware	Battery voltage	`Station ID`	`DWR`	Battery	`Irregular INST`	Limited data are available. This data type allows remote system maintenance checks.

Hardware data types are not commonly available have been implemented as a test and to allow for greater future use.

HydroBase Time Series Types and Standard Time Series Identifier Fields (Reservoir Data)
Reservoir Group Table 1 of 2

Data Group	Data Type Description	Location	Data Source	Data Type	Available Intervals and Time Scale	Comments
Reservoir	Field Measurements	WDID	DWR, other	ResMeasElev, ResMeasEvap, ResMeasFill, ResMeas Release, ResMeas Storage Old (obsolete) data type was RSTO.	Day INST, Day ACCM, Day ACCM, Day ACCM, Day ACCM	Reservoir measurements are often recorded at the beginning or end of the month.
	Pool Elevation	Station ID or State of CO Abbrev.	DWR, other	PoolElev	Irregular INST	Real-time data for reservoirs are recorded using a station abbreviation that does not match a WDID.
	Release Class (showing water color)	WDID	DWR	RelClass-SFUT	Day MEAN, Month INST or ACCM, Year INST or ACCM	SFUT is encoded as: S:s F:f U:u T:t s = source f = from u = use t = type Annual values are for irrigation year (Nov-Oct).
	Release Comment (the acreage for a release and string data flag)	WDID	DWR	RelComment	Year INST or ACCM	See DivComment comments. Sometimes acreage is associated with reservoirs. Annual values are for irrigation year (Nov-Oct).
	Release Total (sum of all RelClass records for a structure).	WDID	DWR	RelTotal	Day MEAN, Month INST or ACCM, Year INST or ACCM	Annual values are for irrigation year (Nov – Oct).

HydroBase Time Series Types and Standard Time Series Identifier Fields (Reservoir Data)
Reservoir Group Table 2 of 2

Data Group	Data Type Description	Location	Data Source	Data Type	Available Intervals and Time Scale	Comments
Reservoir	Release (instantaneous)	Station ID	DWR, other	Release	Irregular INST	Real-time data for reservoirs are recorded using a station abbreviation that does not match a WDID.
	Reservoir Storage (end of month).	WDID	USBR, DWR, other	ResEOM Old (obsolete) data type was RSTO.	Month INST	Few time series are available.
	Reservoir Storage (end of year).	WDID	USBR, DWR, other	ResEOY	Year INST	From *annual_res* table. Annual value is for irrigation year (Nov-Oct).
	Storage (instantaneous)	Station ID or State of CO Abbrev.	DWR, other	Storage	Irregular INST	Real-time data for reservoirs are recorded using a station abbreviation that does not match a WDID.

HydroBase Time Series Types and Standard Time Series Identifier Fields (Stream Data)

Data Group	Data Type Description	Location	Data Source	Data Type	Available Intervals and Time Scale	Comments
Stream	Natural Flow	Station ID	USBR	NaturalFlow Old (obsolete) data type was Nat_flow, NQME	Month INST or ACCM	
	Stage	Station ID	DWR, other	Stage	Irregular INST	Real-time data.
	Streamflow	DWR Abbrev. or USGS station ID	DWR, USGS, other	Streamflow Old (obsolete) daily, monthly data type was QME. Old real-time data type used RT_rate and scenario DISCHRG or other *VAXfield* to indicate channel.	Day MEAN, Month INST or ACCM, Irregular INST	Real-time data use Irregular time interval.
	Streamflow (maximum of daily mean)	Station ID	DWR, USGS	StreamflowMax Old (obsolete) data type was Maxq, Maxflow.	Month INST	
	Streamflow (minimum of daily mean)	Station ID	DWR, USGS	StreamflowMin Old (obsolete) data type was Minq, Minflow.	Month INST	
	Water temperature (instantaneous)	Station ID, State of CO Abbrev.	DWR, other	WatTemp	Irregular INST	Real-time data, using identifier that does not match USGS or other identifier for historical data.

HydroBase Time Series Types and Standard Time Series Identifier Fields (Water Information Sheet Data)

Data Group	Data Type Description	Location	Data Source	Data Type	Available Interval and Time Scale	Comments
WIS	Water Information Sheet (WIS) cell values, over time	WIS row identifier. For example, structures have an identifier wdid:NNNNNNN, where the leading "wdid:" is a literal string and the following information is the actual WDID. Similarly, stations start with "stat:", followed by a station ID; confluences with "conf:", followed by the HydroBase *wd_water* numbers for the tributary and the larger stream; other row types with "othr:", followed by a sequential number in the WIS.	`DWR`	Data types match the WIS columns, as follows: `WISPointFlow`, `WISNaturalFlow`, `WISDeliveryFlow`, `WISGainLoss`, `WISRelease`, `WISPriorityDiversion`, `WISDeliveryDiversion`, `WISTribNaturalFlow`, `WISTribDeliveryFlow`, `WISDryRiver` (not currently implemented – may be implemented as a data flag in the future).	`Day MEAN`	The scenario part of the time series identifier is set to the sheet name. Over time, WIS with a particular sheet name may be modified in format. The combination of sheet name and row identifier can be used to find data. The time series description is set to the row label. Data values are as stored for the WIS, which reflect the gain method used when the sheet was stored.

HydroBase Time Series Types and Standard Time Series Identifier Fields (Well Data)

Data Group	Data Type Description	Location	Data Source*	Data Type	Available Intervals and Time Scale	Comments
Well	Well level (elevation)	Location identifier, based on the current data source. For example, if the data source is USGS, the location identifier will be the USGS identifier.	BJORKLUND CH2MHILL CSU CWSD DWR FOX HALAPASKA HILLIER MCCONAGHY NELSON ROBSON ROBSONBANT SCHNEIDER SEO SMITH SOUTHMETRO SPDSS USGS USGS_NAWQA WILSON *as of 2005-06-16	WellLevel	Day INST, Irregular INST	Daily data are historical measurements, often at the ends of a month. A well may have multiple identifiers. However, the identifier presented in TSTool is that corresponding to the current data source. Use StateView to see alternate identifiers for the location, to cross-reference with data outside of HydroBase. Irregular data are real-time using state station abbreviations, which do not match the identifier for historical data.

Limitations

HydroBase has the following limitations related to time series storage:

- The station and structure measurement types and time series tables defined in HydroBase do not always allow information to be determined from database records. Instead, some time series properties must be hard-coded based on the table design. For example, the *meas_type* table has a `MeanTemp`, `MaxTemp`, `MinTemp` types defined, but these refer primarily to the separate daily tables for such data. The *monthly_temp* table includes *avg_max_t*, *avg_min_t*, and *mean_t* fields that do not correspond one-to-one with *meas_type* values. Therefore, applications like TSTool use data types that are not specifically defined as strings in HydroBase, which have consequently been hard-coded. This is an issue with station and structure time series.
- Real-time data types in HydroBase do not directly translate to time series data types used in TSTool. An effort has been made to be as consistent as possible while using data types that can be understood by users.
- Data units are not defined consistently in tables. Some tables have a units string and others do not and the units abbreviations are not always consistent (units of "A" are often used for acre-feet and "C" for CFS). A master units table is not used in HydroBase to enforce data units consistency throughout the database.
- The time scale for time series (whether accumulated, instantaneous, or mean) is not automatically determined from the data type and interval. Users much understand how to interpret the data, in particular when changing the data interval.