A Community Data Model for Hydrologic Observations

Jeffery S. Horsburgh[1], David G. Tarboton[1] and David R. Maidment[2]

Paper prepared for presentation at the CUAHSI Hydrologic Information System Workshop, Duke University, Durham, North Carolina, July 2005

Abstract

The CUAHSI Hydrologic Information System project is developing information technology infrastructure to support hydrologic science. Part of this includes a data model for the storage and retrieval of hydrologic observations in a relational database. The purpose for a hydrologic observations database is to store hydrologic observations data in a system designed to facilitate data retrieval for integrated analysis of information collected by multiple investigators. It is intended to provide a standard format to facilitate the effective sharing of information between investigators and to facilitate analysis of information within a single study area or hydrologic observatory, or across hydrologic observatories and regions. The hydrologic observations data model is designed to store hydrologic observations and sufficient ancillary information (metadata) about the observations to allow them to be unambiguously interpreted and used and provide traceable heritage from raw measurements to usable information. A relational database format is used to provide querying capability to facilitate data retrieval in support of a diverse range of analyses. An initial data model design was presented at the CUAHSI Hydrologic Information System Workshop held in Austin during March, 2005. An independent review of this initial design identified significant issues that needed to be addressed. This paper presents a redesign of this data model that addresses these issues, to the extent possible within the scope of a relational database model, for the storage and retrieval of point observations.

Introduction

The Consortium of Universities for the Advancement of Hydrologic Science, Inc. (CUAHSI) is an organization representing more than 100 universities sponsored by the National Science Foundation to provide infrastructure and services to advance the development of hydrologic science and education in the United States. The CUAHSI Hydrologic Information System (HIS) project's purpose is to improve infrastructure and services for hydrologic information acquisition and analysis. The project is examining how hydrologic data can be better assembled and analyzed to support hydrologic science and education. As presently conceived, the CUAHSI Hydrologic Information System has four components (Figure 1):

- a *Hydrologic Observations Database*, which is a relational database containing observational data on streamflow, climate, water quality, groundwater levels, and other data measured at monitoring points;

[1] Utah Water Research Laboratory, Utah State University
[2] Center for Research in Water Resources, University of Texas at Austin

- A *Digital Watershed*, which synthesizes the Hydrologic Observations Database with GIS data, weather and climate grids and remote sensing data to form a comprehensive depiction of the water environment of a hydrologic region;
- A *Hydrologic Analysis System*, which supports analysis of fluxes, flow paths, residence times, and mass balances on the Digital Watershed;
- A *Hydrologic Digital Library*, which stores and provides internet access to digital products from all parts of the Hydrologic Information System.

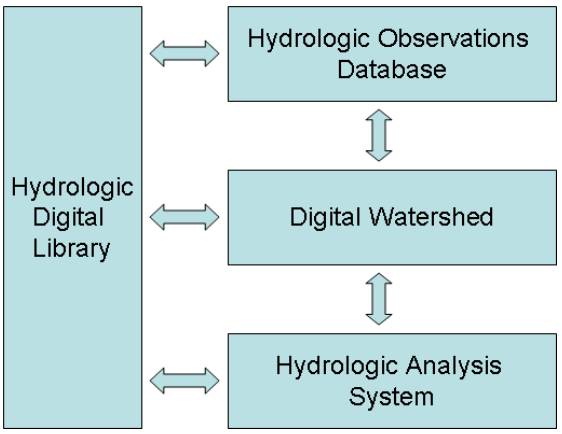


Figure 1. CUAHSI Hydrologic Information System Components

The hydrologic observations data model is the template for the Hydrologic Observations Database and is designed to store hydrologic observations and sufficient ancillary information (metadata) about the observations to allow them to be unambiguously interpreted and used. The metadata will also provide traceable heritage from raw measurements to usable information. A relational database format is used to provide querying capability that facilitates data retrieval in support of a diverse range of analyses. Reliance on databases, and tables within databases also provides the capability to have the model scalable from the observations of a single investigator in a single project, through the multiple investigator communities associated with a hydrologic observatory ultimately to the entire set of observations available to the CUAHSI community.

The hydrologic observations data model is focused on hydrologic observations made at a point. Remotely sensed image or grid data is explicitly excluded as it is handled separately as part of a digital watershed distinct from the hydrologic observations database. Furthermore, information synthesized or derived from raw observations is also excluded, except for simple transformations essential to get the data into a useable form, such as conversions from water level to discharge through a rating curve at a stream gage, transformations from measured voltage to a physical quantity at a probe or instrument, or aggregations from high frequency observations to a desired time step. Synthesis and the derivation of other information and products from hydrologic observations is the role of the Hydrologic Analysis System.

Hydrologic Observations

Many organizations and individuals measure hydrologic variables such as streamflow, water quality, groundwater levels, and precipitation. National databases such as USGS' National Water Information System (NWIS) and USEPA's data Storage and Retrieval (STORET) system contain a wealth of data, but, in general, these national data repositories have different data formats, storage, and retrieval systems, and combining data from disparate sources can be difficult. The problem is compounded when individual investigators are involved (as would be the case at proposed CUAHSI Hydrologic Observatories) because everyone has their own way of storing and manipulating observational data. There is a need within the hydrologic community

for an observations database structure that presents observations from many different sources and of many different types in a consistent format.

Hydrologic observations are identified by the following fundamental characteristics:

- The location at which the observations were made (space)
- The data and time at which the observations were made (time)
- The type of variable that was observed, such as streamflow, water surface elevation, water quality concentration, etc. (variable)

These three fundamental characteristics have been represented by Maidment (2005) as a data cube (Figure 2), where a particular observed data value (D) is located as a function of where it was observed (L), its time of observation (T), and what kind of variable it is (V), thus forming D(L,T,V).

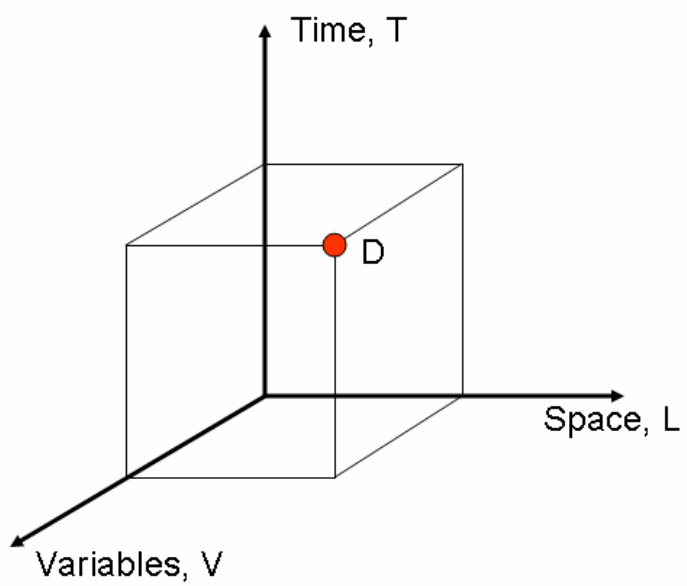


Figure 2. A measured value (D) is indexed by its spatial location (L), its time of measurement (T), and what kind of variable it is (V) (Maidment, 2005).

In addition to these fundamental characteristics, however, there may be many other distinguishing attributes that accompany the observational data. Many of these secondary attributes provide more information about the three fundamental characteristics mentioned above. For example, the location of an observation can be expressed as a text string (i.e., "Bear River Near Logan") or as latitude and longitude coordinates that accurately delineate the location of the observation. Other attributes can provide important context in interpreting the observational data. These include data qualifying comments and information about the organization that collected the data. One of the fundamental design decisions associated with the HOD is how much supporting information to include in the database. This will be discussed in further detail in subsequent sections of this paper.

The ArcHydro Time Series Data Model

In March of 2005, the ArcHydro Time Series Data Model was proposed as a starting point for the HIS HOD structure (Maidment, 2005). This was closely modeled after the time series data model used in ArcHydro (Maidment, 2002). An independent review of this design was undertaken to evaluate whether the ArcHydro Time Series Data Model is adequate to meet the needs of the CUAHSI community and serve as the HIS HOD structure (Tarboton, 2005). Review comments and input were widely requested from scientists familiar with the CUAHSI HIS, from CUAHSI hydrologic observatory planning groups as potential users of the HIS, and from others knowledgeable in database design and dissemination of data. A total of 22 individual sets of review comments were received, and in general the respondents believed that the ArcHydro Time Series Data Model was a good starting point, but that it fell short of providing adequate information to serve as the CUAHSI HIS HOD structure. In addition to comments about the organization and content of the tables in the database, the following is a summary of some of the most important comments and observations that were received as part of the review:

1. In the ArcHydro Time Series Data Model, there is inadequate information to identify the source, heritage, or provenance and give exact definition of the data.
2. The ArcHydro Time Series Data Model does not provide enough information to fully spatially locate a measurement.
3. It is important that the scale of the measurements, defined in terms of their support (averaging domain), spacing, and extent be quantified and associated with measurements.
4. The ArcHydro Time Series data model does not include depth or vertical offset information associated with observations.
5. The ArcHydro Time Series Data Model does not account for censored observations.
6. The classification of time series data types needs to be extended and modified to provide information that guides appropriate interpretation of the data, such as whether the measurements are continuous so that operations such as aggregation or interpolation are meaningful.
7. The focus of the ArcHydro Time Series Data Model on a favored set of proprietary software raised concerns with some reviewers.
8. The ArcHydro Time Series Data model does not include an indication of the quality of the data.

Many of the observations and comments from the review dealt with the general absence of secondary descriptive attributes associated with hydrologic observations within the ArcHydro Time Series Data Model. In order to address these issues and in an effort to meet the needs of the hydrologic community and the CUAHSI HIS for an adequate HOD structure, we have explored alternatives to the ArcHydro Time Series Data Model.

Design Considerations for a Hydrologic Observations Database

In developing a revised HOD structure, we began by extracting from the review comments the design considerations that were considered important by the reviewers. These considerations are:

1. The design should be generic and not rely on unique capabilities of proprietary software. It should be possible to implement the hydrologic observations database in a variety of relational database management systems, including Microsoft Access, Microsoft SQL Server, MySQL, Postgres, and others.
2. The hydrologic observations database should contain at a minimum the important information identified in the reviews of the ArcHydro Time Series Data Model (refer to the section above and to the review document, Tarboton, 2005)
3. The hydrologic observations database should be intuitive enough that users can understand how the data is stored and how to get data into and out of the database.
4. The hydrologic observations database should be capable of storing all information needed to populate a Time Series Object for interfacing with client software designed to view, manipulate, or analyze the data stored in the database.
5. Since the HOD will be the repository for hydrologic observations collected within the proposed hydrologic observatories, it is important that the database be capable of storing not only observations collected by researchers within the observatories, but in addition the HOD should be capable of storing data from the national databases and data collected by state and local agencies or other sources.

These considerations were used in the redesign of the HIS HOD structure.

Alternative Structures

In considering a revised database structure, we asked: **What are the basic attributes to be associated with each single observation and how can these best be organized?** The responses from the review of the originally proposed data model have provided a list of the important attributes to include in the database; however, fundamentally different database structures result from the choice of how much information to associate directly with each observation at the level of a single record, versus how much information is common to a set of observations and can be stored in a linked table. This consideration is important because the structure, number, and nesting of linked tables dictate the efficiency and ease of understanding and use of the data model.

In table 1 we list the attributes associated with each observation that were considered by the reviewers of the originally proposed data model to be necessary parts of the HOD structure. We have attempted to rank these attributes according to how closely they should be associated with the observation value itself, with the presumption that attributes closely associated with the observation value should be stored in the primary observations table while less closely associated information that is common over larger groups of observations should be stored in tables linked to the primary observations table.

Table 1. Ranking of attributes associated with an observation

Attribute	Notes
Value	The observation itself
DateTime	The date and time of the observation (including time zone in which it occurred or offset relative to UTC)
Variable	The physical quantity that the value is measuring (e.g. streamflow, precipitation, water quality)
Location	The location of the observation (i.e., latitude and longitude)
Units	The units (e.g. m or m^3/s) and unit type (e.g. length or volume/time) associated with the variable
Interval	The interval over which the observations were collected or implicitly averaged by the measurement method and whether the observations are regularly recorded on that interval
Offset	Distance from a reference point to the location at which the observation was made (e.g., 5 meters below water surface)
OffsetType/ Reference Point	The reference point from which the offset to the measurement location was measured (i.e., water surface, stream bank, snow surface)
Data Type	An indication of the kind of quantity being measured (e.g., an instantaneous or cumulative measurement)
Organization	The organization or entity providing the measurement
Censoring	An indication of whether the observations is censored or not
Data Qualifying Comments	Comments accompanying the data that can affect the way the data is used or interpreted (e.g., holding time exceeded, sample contaminated, provisional data subject to change, etc.)
Analysis Procedure	An indication of what method was used to collect the observation (e.g., dissolved oxygen by field probe or dissolved oxygen by Winkler Titration)
QA/QC	An indication of the quality of the data
Source Database	An indication of the original source of the observation (e.g., USGS NWIS, EPA STORET, local investigator, etc.)
Sample Medium	The medium in which the sample was collected (e.g., water, air, sediment, etc.)
Value Type	An indication of whether the value represents an actual measurement, a calculated value, or is the result of a model simulation

Two fundamentally different database structures were proposed by two different reviewers of the original data model. To evaluate the impact that these different designs have on the characteristics of the observations database, we populated the two different structures with a single dataset. For this comparison, the designs were modified from what the reviewers had suggested so that they both contained the same fields (data attributes), but differed in the way that the tables were organized.

The first structure is very similar to the ArcHydro Time Series Data Model, but it attempts to include much of the additional information requested by many of the reviewers. For the

purposes of this example, we considered the first proposed structure to be inclusive of the ArcHydro Time Series Data Model. The second proposed structure is fundamentally different from the first proposed structure in that it stores much of the metadata associated with the observations in a linked table rather than in the same table as the observations themselves.

The following figures illustrate the main differences between the two structures proposed by the reviewers. Structure 1 (Figure 3) proposes direct inclusion of a larger amount of ancillary information as record level metadata in the time series table through identifiers that link in to adjoining tables. Structure 2 (Figure 4) proposes that all metadata information should be referenced through one TSType table linked to the main time series table, with other information linked to the TSType table. The remaining tables were identical in both databases. The first design is intended to facilitate querying directly based on a wide range of attributes at the cost of storing a number of metadata identifiers with each observation. The second design minimizes the number of metadata identifiers to be stored with each observation with the intent of reducing the size of primary time series table, but at the expense of a larger TSType table because there are more unique "type" combinations.

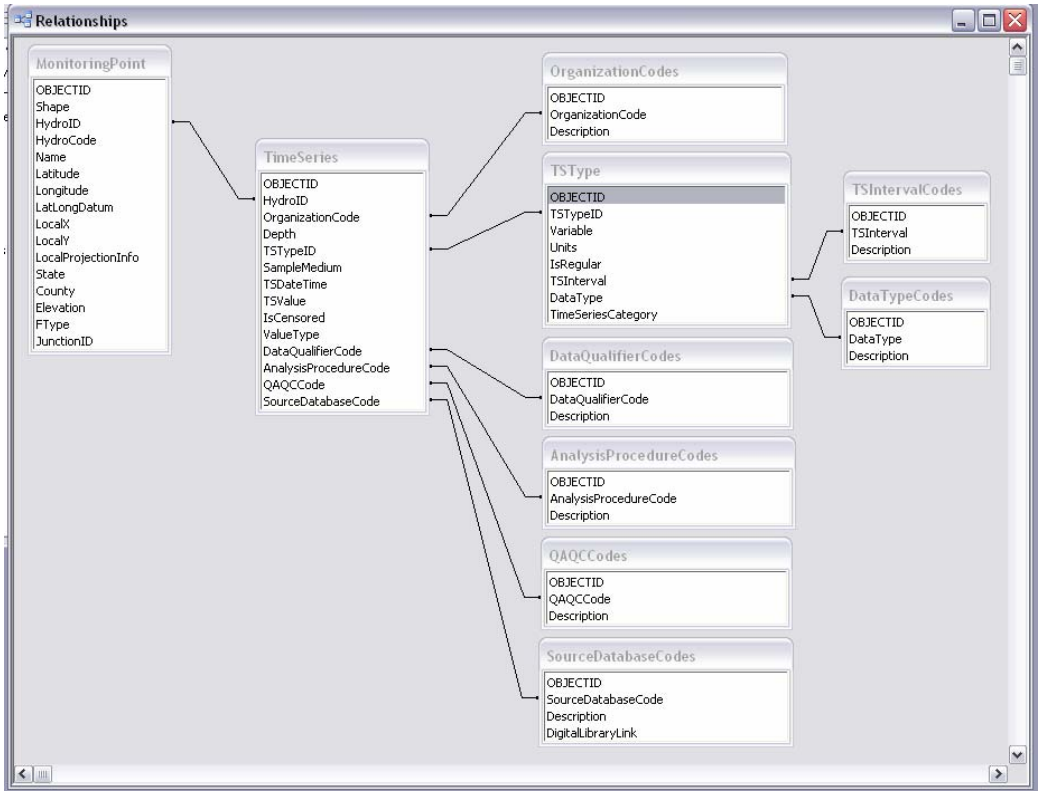


Figure 3. Hydrologic Observations Database Alternative Structure 1.

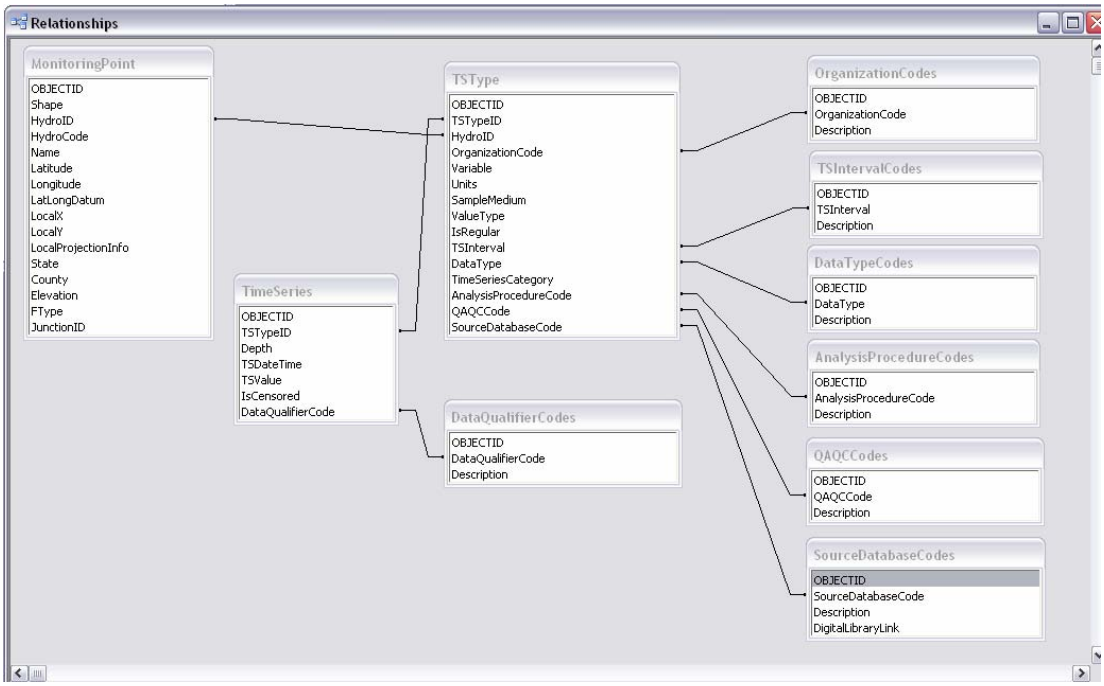


Figure 4. Hydrologic Observations Database Alternative Structure 2.

Both of these proposed structures were considered to be viable designs for the HIS HOD structure, and were, therefore, considered in the redesign of the HIS HOD database structure. It was anticipated, however, that each of these proposed structures would have implications and tradeoffs with regard to the design decisions listed above, and so a series of simple tests were performed to evaluate the two proposed structures. These tests are described in the following section.

Alternative Structure Tests

The two structures described in the previous section were evaluated through a series of simple tests that were designed to provide information about which of the structures was more appropriate to serve as the HIS HOD structure. Both database structures were implemented in Microsoft Access and were populated with USGS water quality data for a single 8-digit HUC (16010203 – Little Bear-Logan[3]). At the time it was downloaded, this dataset included 127 monitoring points, 369 different water quality variables, and 11,885 individual water quality observations.

All of the tables in the two databases are exactly the same, except for the TimeSeries and the TSType tables. In both databases, the TimeSeries table contains 11,885 records (one for each observation), but the TimeSeries table in proposed structure 1 contains metadata information that has been moved to the TSType table in proposed structure 2. The result of this fundamental difference is that the TSType table in proposed structure 1 contains 369 records (one for each unique variable), but the TSType table in proposed structure 2 contains 4359 records (one for each unique combination of location, organization, variable, units, sample medium, value type,

[3] http://nwis.waterdata.usgs.gov/ut/nwis/qwdata?huc_cd=16010203&format=rdb

etc.). In the context of the HOD database, it should be noted that the number of records in the TSTypes table of structure 2 could increase dramatically as more locations, organizations, variables, units, etc. are added to the database.

In terms of size on disk, proposed structure one is approximately 2.3 MB in size, and proposed structure 2 is approximately 6 MB in size. It is anticipated that structure 1 would be smaller than structure 2 as long as there is a relatively small number of observations (records in the TimeSeries table) and a relatively large number of variables (records in the TSType table). Conversely, it is anticipated that structure 1 would likely be larger than structure 2 if there were many observations (records in the TimeSeries table), but few variables (records in the TSType table). No tests were performed to confirm these observations.

Another simple test involved creating a simple query to retrieve data from the databases. This simple query test was not intended to demonstrate completely the differences in querying information out of the two databases. Rather, it is used here to demonstrate what is perhaps one of the most important differences between the two alternative structures. The following query was created so that it could be tested in both databases:

> *"Give me a list of the HydroID, HydroCode, and Name of all sampling locations at which water temperature data has been collected."*

Structure 1 allows the user to create a query to return the requested information by specifying criteria on the TSTypeID field *or* the Variable field to retrieve the requested information. The following are SQL statements used to return the requested information:

```
SELECT DISTINCT MonitoringPoint.HydroID, MonitoringPoint.HydroCode, MonitoringPoint.Name,
TimeSeries.TSTypeID
FROM MonitoringPoint INNER JOIN TimeSeries ON MonitoringPoint.HydroID = TimeSeries.HydroID
WHERE (((TimeSeries.TSTypeID)=10));
```

OR

```
SELECT DISTINCT MonitoringPoint.HydroID, MonitoringPoint.HydroCode, MonitoringPoint.Name,
TSType.Variable
FROM (MonitoringPoint INNER JOIN TimeSeries ON MonitoringPoint.HydroID = TimeSeries.HydroID) INNER
JOIN TSType ON TimeSeries.TSTypeID = TSType.TSTypeID
WHERE (((TSType.Variable) Like "Temperature, water*"));
```

Since there are many records in the TSType table of Structure 2 where the variable is "Temperature, water", this limits the ability to query in that we must specify criteria on the Variable field unless we know all of the TSTypeIDs where the variable is equal to "Temperature, water." The following is the query executed on structure 2 to return the requested information

```
SELECT DISTINCT MonitoringPoint.HydroID, MonitoringPoint.HydroCode, MonitoringPoint.Name,
TSType.Variable
FROM MonitoringPoint INNER JOIN TSType ON MonitoringPoint.HydroID = TSType.HydroID
WHERE (((TSType.Variable) Like "Temperature, water*"));
```

The queries to both database structures are nearly the same, but the criteria (**bold**) are different. In structure 1, we can use TSTypeID = 10 to return water temperature because 10 as the

TSTypeID for water temperature is unique. In Structure 1 we can also put criteria on the variable name because it is unique (i.e., we can do either to return the same information). In structure 2, there are many TSTypeIDs that represent water temperature, so we can only put criteria on the variable name unless we know all of the integer TSTypeIDs that correspond to water temperature (there are 112 of them). It is important to consider that to put criteria on the variable name we must deal with the vocabulary of the Variable issue (i.e., is it "Temperature, water, degrees Celsius" or "Water Temperature, degrees Celsius" or "Water Temperature, deg. C," etc. This can be controlled to some degree through the use of a controlled vocabulary in the variable field.

Revised Hydrologic Observations Database Structure Design

After evaluating the two proposed database structures, we have settled on a structure that falls somewhere in between the two. In general, we preferred structure 1 because it was easier to populate and more intuitive to query. However some changes have been made to proposed structure 1 to meet the needs of the CUAHSI HIS and to address the comments from the reviewers. For starters, some of the metadata will be maintained at the record level in the Observations table (formerly the TimeSeries table), and, where appropriate, some has been moved to the ObservationTypes table (formerly the TSType table). This will avoid what we perceive to be unnecessary duplication in the ObservationTypes table, and it will make it easier to retrieve data from the database based on a variable type. In addition, we have changed the names of some of the tables and fields to reflect that the database is storing hydrologic observations. Figure 5 shows the table schema for the revised HIS hydrologic observations database. Appendix A provides a data dictionary that lists the tables in the database, the names and data types of each of the fields in the tables, and provides a description of the information contained in each of the fields.

In addition to the changes listed in the preceding paragraph, we have made several other modifications to the database structure so that it differs from those that were tested. They are as follows:

1. We have added an ObservationsCatalog table to the database. Although not required to maintain the integrity of the data, this table provides a listing of all of the monitoring point and observation type combinations in the database. This provides a means by which a user can get simple descriptive information about the variables observed at a location, the most common anticipated query, without the overhead of querying the entire time series table, which can become quite large.
2. We have added a UTCOffset field to the Observations table to ensure that local times recorded in the database can be referenced to standard time and to enable comparison of results across databases that may store observations collected in different time zones (i.e., compare observations from one hydrologic observatory to those collected at another hydrologic observatory located across the country). A design choice here was to have UTCOffset as a record level qualifier because even though the time zone and hence offset is likely the same for all measurements at a monitoring point, the offset changes due to daylight savings. Some investigators may run data loggers on standard time, while others may adjust for daylight saving or use universal time. To avoid the necessity to keep track

of the system used, or impose a system that might be cumbersome and lead to errors we decided that if the offset was always recorded the precise time would be unambiguous and would reduce the chance for interpretation errors.

3. We have added an ObservationID to the Observations table to uniquely identify each individual observation and serve as an identifier for use in the definition of logical groupings of observations and sets of observations used to derive other observations.
4. We have added two tables, ObservationGroups and GroupDescriptions, which enable the logical grouping of observations (i.e., assigning all observations from a single reservoir profile to one group). These tables provide a means of grouping together observations that are logically related.
5. We have added a DerivedFromID to the Observations table and a DerivedFrom table to the database. The DerivedFromID points to the DerivedFrom table where the observations from which a quantity was derived are listed (e.g. a daily average discharge value could be linked to the 96 15 minute unit values from which it was derived, or a snow water equivalent value could be linked to the depth and density values from which it was derived).
6. We have combined the AnalysisProcedureCodes and QAQCCodes tables into a single table that indicates the method used to collect the observation and the QAQC associated with that method. The description field in this table would describe both the analysis procedure and the QAQC level.
7. We have converted the DataType field to a text field with a controlled vocabulary (rather than a coded value domain) eliminating the need for a value coding table. We have added some additional categories to the DataTypes and have renamed the TSInterval field as ObservationSupport to use this field to specifically quantify the time support scale of the measurements. The definitions of DataTypes and support scale are given below.
8. We have added a CategoryDefinitions table that stores the categories associated with categorical observations. These observations are encoded as double values in the Observations table.

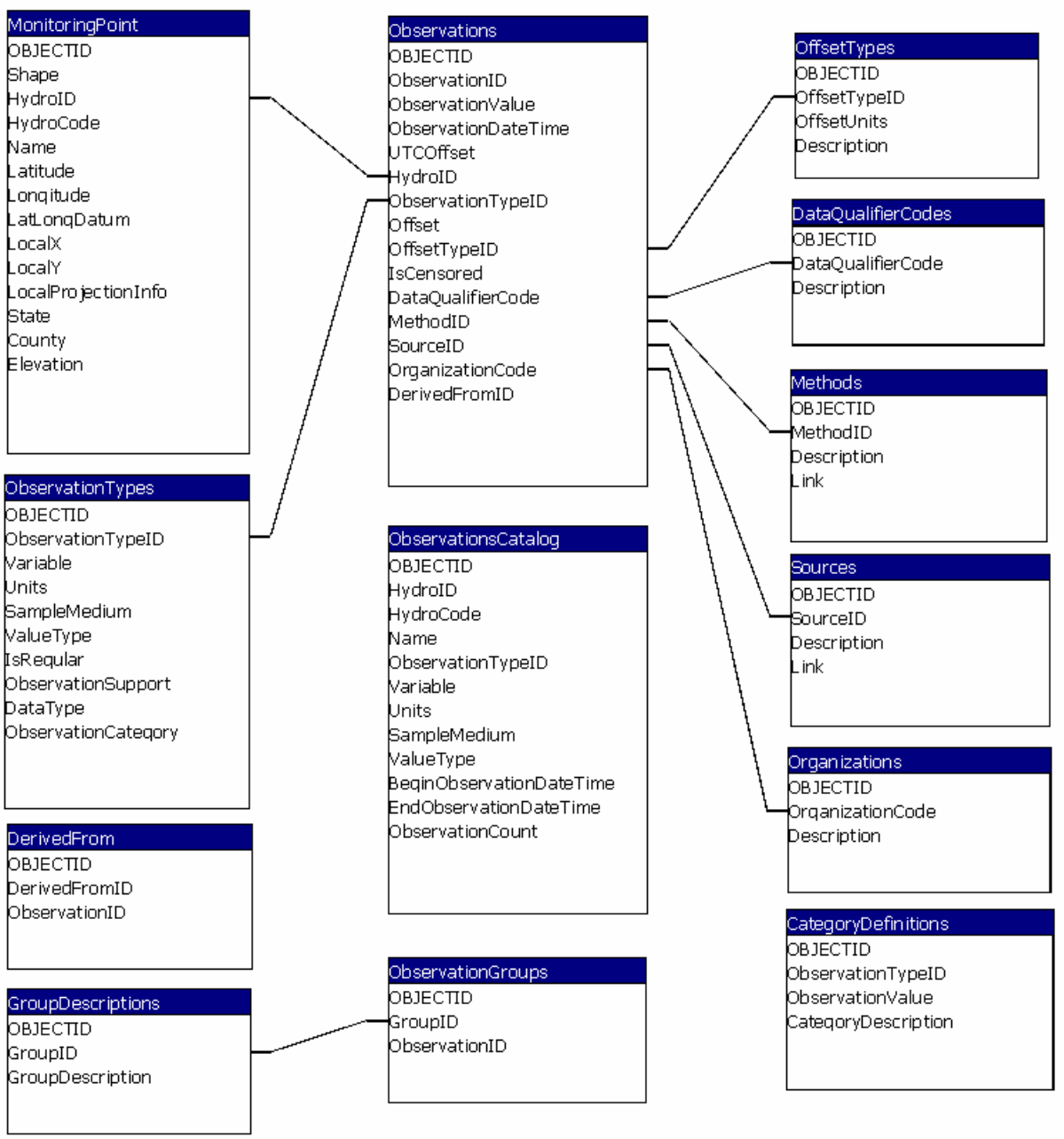


Figure 5. Proposed Hydrologic Observations Database Structure.

DataType and Support Scale

In interpreting observations that comprise a time series it is important to know the scale information associated with the observations. Blöschl and Sivapalan (1995) review the important issues. Any set of observations is quantified by a scale triplet comprising support, spacing and extent, illustrated in Figure 6.

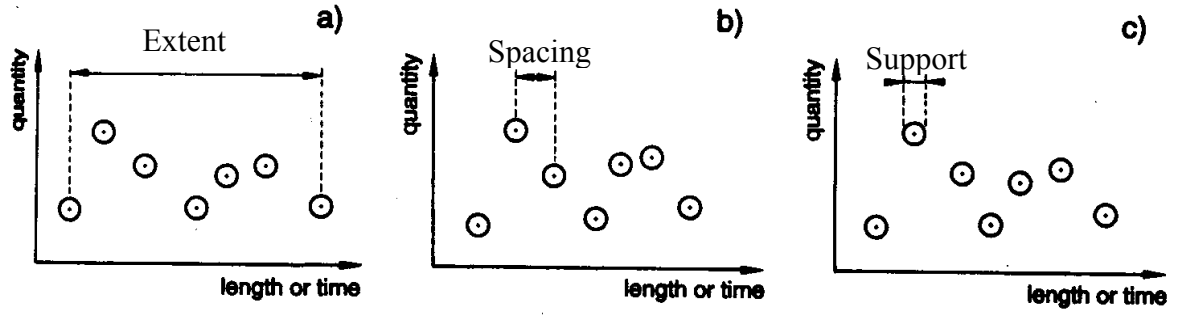


Figure 6. The Scale Triplet of Measurements (a) Extent, (b) Spacing, (c) Support. (from Blöschl, 1996)

Extent is the full range over which the measurements occur, spacing is the spacing between measurements and support is the averaging interval or footprint implicit in any measurement. In the proposed Hydrologic Observations Data model extent and spacing are properties of multiple measurements and are defined by the DateTime associated with observations. Instead of a variable TSinterval that was in the preliminary data model we have included a field called ObservationSupport in the time series table to explicitly quantify support. Figure 7 shows some of the implications associated with support, spacing and extent in the interpretation of time series observations.

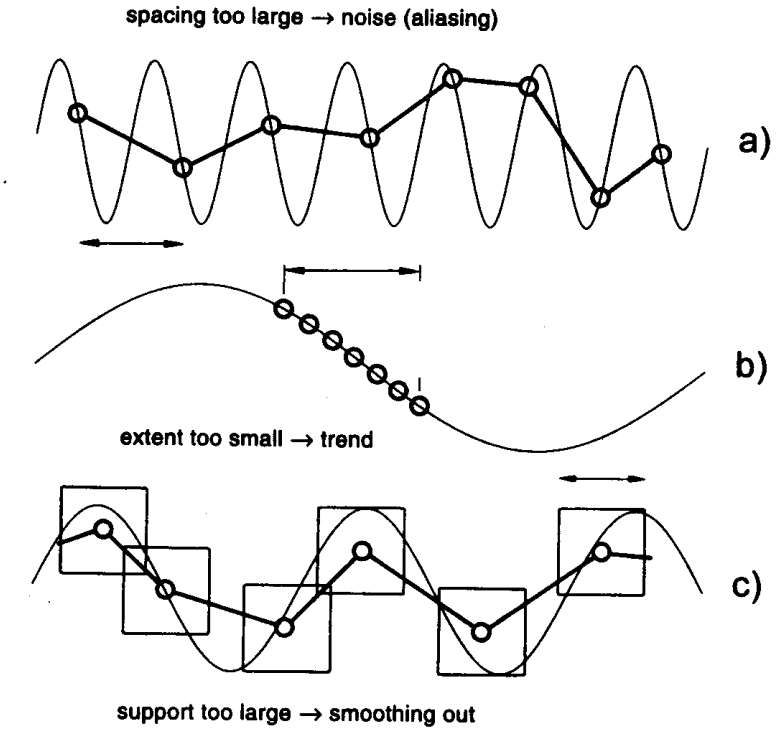


Figure 7. The effect of sampling for measurement scales not commensurate with the process scale. (a) Spacings larger than the process scale cause aliasing in the data; (b) Extents smaller than the process scale cause a trend in the data; (c) Supports larger than the process scale cause excessive smoothing in the data. (from Blöschl, 1996)

In the proposed Hydrologic Observations Data model the following data types are suggested. These are extensions from the initial ArcHydro time series data model.

1. *Continuous* data – the phenomenon, such as streamflow, Q(t) is specified at a particular instant in time and measured with sufficient frequency (small spacing) to be interpreted as a continuous record of the phenomenon.
2. *Instantaneous* data – the phenomenon is sampled at a particular instant in time but with a frequency that is too coarse for interpreting the record as continuous. This would be the case when the spacing is significantly larger than the support and the time scale of fluctuation of the phenomenon, such as for example infrequent water quality samples.
3. *Cumulative* data – the data represents the cumulative value of a variable measured or calculated up to a given instant of time, such as cumulative volume of flow or cumulative precipitation: $V(t) = \int_0^t Q(\tau)d\tau$, where τ represents time in the integration over the interval [0,t]. To unambiguously interpret cumulative data one needs to know the time origin. We suggest the convention of using a cumulative record with an ObservationValue of zero to initialize or reset cumulative data. With this convention cumulative data should be interpreted as the accumulation over the time interval between the DateTime of the zero record and the current record at the same observation position. Observation position is defined by a unique combination of HydroID, ObservationType, Offset and OffsetType. All four of these quantities comprise the unambiguous description of the position of an observation and there may be multiple time series associated with multiple observation positions (e.g. redundant rain gauges with different offsets) at a location.
4. *Incremental* data – the value represents the incremental value of a variable over a time interval Δt such as the incremental volume of flow, or incremental precipitation: $\Delta V(t) = \int_{t-\Delta t}^t Q(\tau)d\tau$. As for cumulative data, unambiguous interpretation requires knowledge of the time increment. Here we suggest the convention of using ObservationSupport if this is given, or the time interval from the previous observation at the same position if ObservationSupport is not given or is 0. This accommodates incremental type precipitation data that is only reported when the value is non-zero, such as NCDC data.
5. *Average* data – the value represents the average over a time interval, such as daily mean discharge or daily mean temperature: $\overline{Q}(t) = \frac{\Delta V(t)}{\Delta t}$. The averaging interval is quantified by ObservationSupport in the case of regular data (as quantified by the IsRegular field) and by the time interval from the previous observation at the same position for irregular data.
6. *Maximum* data – the value is the maximum value occurring at some time during a time interval, such as annual maximum discharge or a daily maximum air temperature. Again unambiguous interpretation requires knowledge of the time interval. We suggest the convention that the time interval is the ObservationSupport for regular data and the time interval from the previous observation at the same position for irregular data.

7. *Minimum* data – the value is the minimum value occurring at some time during a time interval, such as 7-day low flow for a year, or the daily minimum temperature. The time interval is defined similarly to Maximum data.
8. *Constant over interval* data – the value is a quantity that can be interpreted as constant over the time interval from the previous measurement.
9. *Categorical* data – the value is a categorical rather than continuous valued quantity. Mapping from ObservationValue values to categories is through the CategoryDefinitions table.

Examples

To demonstrate the capability of this design to store a diverse set of hydrologic observations Appendix B gives examples of how an illustrative set of observations would be represented in this database design.

Discussion

This data model design was conceived with a number of considerations in mind, some of which came from the review of the initial data model and others of which emerged during discussion of this design. These are reviewed here to give a sense of some of the capabilities envisaged for the data model.

The DerivedFrom and ObservationGroups table fulfill the function of grouping observations for different purposes. These are tables where the same identifier (DerivedFromID or GroupID) can appear multiple times in the table associated with different ObservationIDs thereby defining the associated group of records. In the DerivedFrom table this is the sole purpose of the table and each group so defined is associated with a record in the Observations table (through the DerivedFromID field in that table). This record would have been derived from the observations identified by the group. The method of derivation would be given through the methods table associated with the observation. This construct is useful for example to identify the 96 15 min unit streamflow values that go into the estimate of the mean daily streamflow. Note that there is no limit as to how many groups an observation may be associated with, and observations that are derived from other observations may themselves belong to groups used to derive other observations (e.g. the daily minimum flow over a month derived from daily observations derived from 15 min unit values). Note also that a derived from group may have as few as one observation for the case where an observation is derived from a single more primitive observation (e.g. Discharge from Stage). Through this construct the data model has the capability to store raw observations and simple derivatives preserving the connection of each observation to its more primitive raw measurement.

In the design presented we have represented categorical or ordinal variables in the same table as continuous valued 'double' variables through a numerical encoding of the categorical observation value as a ‘double’ value. The CategoryDefinitions table then associates, for each observation type an observation value with an associated category definition. This is a somewhat cumbersome construct because real valued 'double' quantities are being used as database keys. We do not see this as a significant shortcoming though because typically, in our judgment, only a

small fraction of hydrologic observations will be categorical. An alternative approach could have been to have a separate Observations table for categorical observations.

The Methods and Sources tables both contain links that we have indicated as either a URL or reference to a file in a digital library. It will be important as the database grows and is used over time to ensure that links or URL's included are stable. An alternative approach to external links is to exploit the capability of modern databases to store as fields within a record entire digital documents, such as an html or xml page, PDF document or raw data file. The capability therefore exists to instead have these links refer to a Documents table that would actually contain this metadata information, instead of housing it in digital library. There is some merit in this because then any data exported in Hydrologic Observations Data model format could take with it the associated metadata required to completely define it as well as the raw data upon which it is derived. This however has the disadvantage of increasing (perhaps substantially) the size of database file containing the data and being distributed to users. The implications of this idea have not been fully explored. It is mentioned here as a possibility worthy of further consideration.

A considerable portion of hydrologic observations data is in the form of time series. This was why the initial model was based on the ArcHydro Time Series Data Model. The proposed design has not specifically highlighted time series capabilities, nevertheless the data model has inherited the key components from the ArcHydro Time Series Data Model to give it time series capability. In particular one observation DataType is "Continuous," designed to indicate that the observations are collected with sufficient frequency as to be interpreted as a smooth time series. The IsRegular field also facilitates time series analysis because certain time series operations (e.g. Fourier Analysis) are predisposed to regularly sampled data. At first glance it may appear that there is redundancy between the Isregular field and the DataType "Continuous" but we chose to keep these separate because there are regularly sampled quantities for which it is not reasonable to interpret the values as "Continuous". For example monthly grab samples of water quality are not continuous, but are better categorized as having DataType, "Instantaneous". Note that the data model does not explicitly store the time interval between measurements, nor does it indicate where a continuous series has data gaps. Both these are required for time series analysis, but are inherently not properties of single measurements. The time interval is the time difference between sequential regular measurements, something that could be easily computed from DateTime values by analysis tools. The inference of measurement gaps (and what to do about them) from DateTime values we also regard as analysis functionality left for the Hydrologic Analysis System to handle.

Conclusions

This paper has presented the design for a community hydrologic observations database structure that is designed to store hydrologic observations in a flexible, relational database system to facilitate data retrieval for integrated analysis of information collected by multiple investigators. The design represents an evolution of the initial ArcHydro time series database design, to address the specific needs of the CUAHSI community identified by reviewers of the initial design. The data model is focused on storing the original observations, simple derived quantities, and ancillary information (metadata) sufficient to allow unambiguous interpretation of

data, while at the same time providing traceable heritage from raw measurements to usable information. It is recommended that this data model be implemented and tested in a number of database systems to fully evaluate its suitability for adoption as a CUAHSI hydrologic observations data model standard.

Acknowledgements

This work is supported by the National Science Foundation under grant #EAR-0413265 to CUAHSI and the University of Texas at Austin. The views and conclusions expressed are those of the authors and do not necessarily reflect the views of the U.S. Government.

References

Blöschl, G. and M. Sivapalan, (1995), "Scale Issues in Hydrological Modelling: A Review," Hydrological Processes, 9(1995): 251-290.

Blöschl, G., (1996), Scale and Scaling in Hydrology, Habilitationsschrift, Weiner Mitteilungen Wasser Abwasser Gewasser, Wien, 346 p.

Maidment, D. R., ed. (2002), Arc Hydro Gis for Water Resources, ESRI Press, Redlands, CA, 203 p.

Maidment, D.R. 2005. A Data Model for Hydrologic Observations. Paper prepared for presentation at the CUAHSI Hydrologic Information Systems Symposium. University of Texas at Austin. March 7, 2005.

Tarboton, D.G. 2005. Review of Proposed CUAHSI Hydrologic Information System Hydrologic Observations Data Model. Utah State University. May 5, 2005.

Appendix A
Table and Field Structure for the Proposed HIS Hydrologic Observations Database

The following is a description of the tables in the proposed hydrologic observations database schema, a listing of the fields contained in each table, a description of the data contained in each field and its type, examples of the information to be stored in each field where appropriate, and any additional information about each field. Values in the example column should not be considered to be inclusive of all potential values, especially in the case of fields that will require a controlled vocabulary. We have developed some suggestions for the controlled vocabulary for some fields, but anticipate that these will need to be extended and adjusted.

Table: CategoryDefinitions

Associates observation value with the definition of a category for categorical variables

Field Name	Data Type	Description	Example	Notes
OBJECTID	Integer (Autonumber)			
ObservationTypeID	Integer	Integer identifier that references the observation type record of a categorical variable		This identifies the specific type of observations for which a value to category mapping applies and avoids conflicts where the same numerical value may map into different categories for different observation types.
ObservationValue	Double	Numeric value of Observation	1.0	Although a real number represented as a double these are associated with categories defined in the CategoryDescription field
CategoryDescription	Text	Definition of categorical variable value	"Cloudy"	

Table: DataQualifierCodes

Lists the full descriptions of the data qualifying comments that accompany the data. This table serves to define the controlled vocabulary of text codes stored in the observations table.

Field Name	Data Type	Description	Example	Notes
OBJECTID	Integer (Autonumber)			
DataQualifierCode	Text	Unique code identifying the data qualifying comment	"H"	The following initial controlled vocabulary is suggested: E – Estimated P – Provisional D – Derived H – Holding time for sample analysis exceeded
Description	Text	Full description or text of the data qualifying comment	"Holding time for sample analysis exceeded"	

Table: DerivedFrom

Table that contains the linkage between derived quantities and the observations that they were derived from.

Field Name	Data Type	Description	Example	Notes
OBJECTID	Integer (Autonumber)			
DerivedFromID	Integer	Unique integer identifying the group of observations from which a quantity is derived		
ObservationID	Integer	Integer identifier referencing observations that comprise a group of observations from which a quantity is derived		This corresponds to ObservationID in the Observations table

Table: GroupDescriptions

Lists the descriptions for each of the observation groups that have been formed.

Field Name	Data Type	Description	Example	Notes
OBJECTID	Integer (Autonumber)			
GroupID	Integer	Unique integer identifier for each group of observations that has been formed		This also references to GroupID in the ObservationGroups table
GroupDescription	Text	Text description of the group	"Echo Reservoir Profile 7/7/2005"	

Table: Methods

Lists the methods used to collect the data and provides an indication of the Quality Assurance and Quality Control procedures associated with each method.

Field Name	Data Type	Description	Example	Notes
OBJECTID	Integer (Autonumber)			
MethodID	Integer	Unique integer ID for each measurement method/QAQC combination		
Description	Text	Text description of each method/QAQC combination.	"Total phosphorus measured using EPA procedure XXX with published QAQC plan"	
Link	Hyperlink	Link to a file in digital library or URL that provides a description of the method		

Table: MonitoringPoint

Provides information giving the spatial location at which observations have been collected.

Field Name	Data Type	Description	Example	Notes
OBJECTID	Integer (Autonumber)			
Shape	Binary Object	ESRI geodatabase shape information		
HydroID	Integer	Unique integer ID for each sampling location		Easier to index and query than the HydroCode, which is text
HydroCode	Text	Unique text identifier for each sampling location	"10109000"	This is redundant with HydroID but is retained to provide a recognizable identifier associated with each location useful for error checking
Name	Text	Full name of sampling location	"LOGAN RIVER ABOVE STATE DAM, NEAR LOGAN,UT"	
Latitude	Double	Latitude in decimal degrees		
Longitude	Double	Longitude in decimal degrees		
LatLongDatum	Text	Datum of latitude and longitude	"NAD 83" "NAD 27"	Controlled Vocabulary
LocalX	Double	Local Projection X coordinate		

Field Name	Data Type	Description	Example	Notes
LocalY	Double	Local Projection Y Coordinate		
LocalProjectionInfo	Text	Information describing local projection	"UTMZone12NAD83"	Controlled Vocabulary
State	Text	Name of state in which the sampling station is located	"Utah"	
County	Text	Name of County in which the sampling station is located	"Cache"	
Elevation_m	Double	Elevation of sampling location (in m)		Meters above sea level

Table: ObservationGroups

Lists the groups of observations that have been created and the observations that are within each observation group.

Field Name	Data Type	Description	Example	Notes
OBJECTID	Integer (Autonumber)			
GroupID	Integer	Unique integer ID for each group of observations that has been formed		
ObservationID	Integer	Integer identifier for each observation that belongs to a group		This corresponds to ObservationID in the Observations table

Table: Observations

Stores the actual hydrologic observations.

Field Name	Data Type	Description	Example	Notes
OBJECTID	Integer (Autonumber)			
ObservationID	Integer	Unique integer identifier for each observation		
ObservationValue	Double	Numeric value of observation		Categorical information is stored as a number with the categories defined in observation type table?
ObservationDateTime	Date/Time	Local date and time at which the observation was made		Represented as: MM/DD/YYYY hh:mm:ss.sss Where MM=Month DD = Day YYYY=Year hh = Hour mm = minutes ss.sss = seconds with milliseconds
UTCOffset	Integer	Offset from UTC time at the sampling location		Number of hours
HydroID	Integer	Integer identifier of the sampling location at which the observation was made		This links observations to their locations in the MonitoringPoint table
ObservationTypeID	Integer	Integer identifier that references the variable that was measured		This links observations to their type in the ObservationTypes table

Field Name	Data Type	Description	Example	Notes
Offset	Double	Distance from a datum or control point at which an observation was made		
OffsetTypeID	Integer	Unique integer identifier that references the type of measurement offset		This links observation offsets to their type in the OffsetTypes table
IsCensored	Text	Text indication of whether the data value is censored		Controlled Vocabulary “gt”=greater than “lt”=less than “nc” or blank=not censored
DataQualifierCode	Text	Text code that indicates a data qualifying comment		These codes are defined in the DataQualifierCodes table
MethodID	Integer	Integer identifier that references the measurement method/QAQC combination associated with the observation		This links observations to their method description in the Methods table
SourceID	Integer	Integer identifier that references the record in the Sources table giving the source of the observation		
OrganizationCode	Text	Unique text code that identifies the organization that colledted the data		The organization table associates the 'short' organization code with a complete organization description

Field Name	Data Type	Description	Example	Notes
DerivedFromID	Integer	Integer identifier for the group of observations that the current observation is derived from		This refers to a group of derived from records in the DerivedFrom table.

Table: ObservationsCatalog

Lists each of the MonitoringPoint/ObservationType combinations in the database as an index to speed some simple queries. This table contains the necessary fields to uniquely identify each sampling location and each measured quantity at that location for the purposes of identifying or displaying what data are available at each location without querying the main Observations table.

Field Name	Data Type	Description	Example	Notes
OBJECTID	Integer (Autonumber)			
HydroID	Integer	Unique integer monitoring point or sampling location identifier		
HydroCode	Text	Unique text identifier for each sampling location		
Name	Text	Full text name of sampling location		
ObservationTypeID	Integer	Integer identifier for each ObservationType		
Variable	Text	Name of the variable corresponding to observation type		
Units	Text	Units of the variable corresponding to observation type		
SampleMedium	Text	The medium of the sample		
ValueType	Text	Text value indicating what type of observation is being recorded		

Field Name	Data Type	Description	Example	Notes
BeginObservationDateTime	Date/Time	Date of the first observation in the series identified by the combination of the HydroID and ObservationTypeID		
EndObservationDateTime	Date/Time	Date of the last observation in the series identified by the combination of the HydroID and ObservationTypeID		
ObservationCount	Integer	The number of observations in the series identified by the combination of the HydroID and the ObservationTypeID		

Table: ObservationTypes

Lists the full descriptive information about what variables have been measured.

Field Name	Data Type	Description	Example	Notes
OBJECTID	Integer (Autonumber)			
ObservationTypeID	Integer	Unique integer identifier for each ObservationType		
Variable	Text	Name of the variable that was measured, observed, modeled, etc.	"Water Temperature"	Controlled Vocabulary
Units	Text	Text units of the observation	"Degrees Celsius"	Controlled Vocabulary
SampleMedium	Text	The medium of the sample	"Surface Water" "Sediment" "Fish Tissue"	Controlled Vocabulary

Field Name	Data Type	Description	Example	Notes
ValueType	Text	Text value indicating what type of observation is being recorded	"Field Observation" "Laboratory Observation" "Model Simulation Results"	Controlled Vocabulary
IsRegular	Boolean	Value that indicates whether the values are from a regularly sampled time series	"True" "False"	Controlled Vocabulary
ObservationSupport	Double	Numerical value in hours that indicates the support (or temporal footprint) for these observations	0, 24	0 is used to indicate a value that is instantaneous. Other values indicate the time over which the observations are implicitly or explicitly averaged
DataType	Text	Text value that identifies the data as one of several types	"Continuous" "Instantaneous" "Cumulative" "Incremental" "Average" "Minimum: "Maximum" "Constant Over Interval" "Categorical"	Controlled Vocabulary
ObservationCategory	Text	General category of the observations	"Climate" "Water Quality" "Groundwater Quality"	Controlled Vocabulary

Table: OffsetTypes

Lists the full descriptive information for each of the measurement offsets.

Field Name	Data Type	Description	Example	Notes
OBJECTID	Integer (Autonumber)			
OffsetTypeID	Integer	Unique integer identifier that identifies the type of measurement offset		
OffsetUnits	Text	Units of the offset distance	"m" for meters	Controlled Vocabulary
Description	Text	Full text description of the offset type	"Below water surface" "Above Ground Level"	Controlled Vocabulary

Table: Organizations

Lists the full descriptive information for each data collection organization.

Field Name	Data Type	Description	Example	Notes
OBJECTID	Integer (Autonumber)			
OrganizationCode	Text	Unique text code that identifies the data collection organization		
Description	Text	Full text description of data collection organizations	"United States Geological Survey"	

Table: Sources

Lists the original sources of the data, including a link to the original data files and metadata that should be contained in the digital library.

Field Name	Data Type	Description	Example	Notes
OBJECTID	Integer (Autonumber)			

Field Name	Data Type	Description	Example	Notes
SourceID	Integer	Unique integer identifier that identifies each data source		
Description	Text	Full text description of the source database	"Text file retrieved from the United States Geological Survey National Water Information System"	
Link	Hyperlink	Link to original data file and associated metadata stored in the digital library or URL of data source		

Appendix B
Examples

The following examples show the capability of the proposed data structure to store different types of hydrologic observations.

Streamflow Stage and Discharge

Both stage measurements and the associated discharge estimates derived from the stage measurements can be stored in the proposed observations database (Figure A.1).

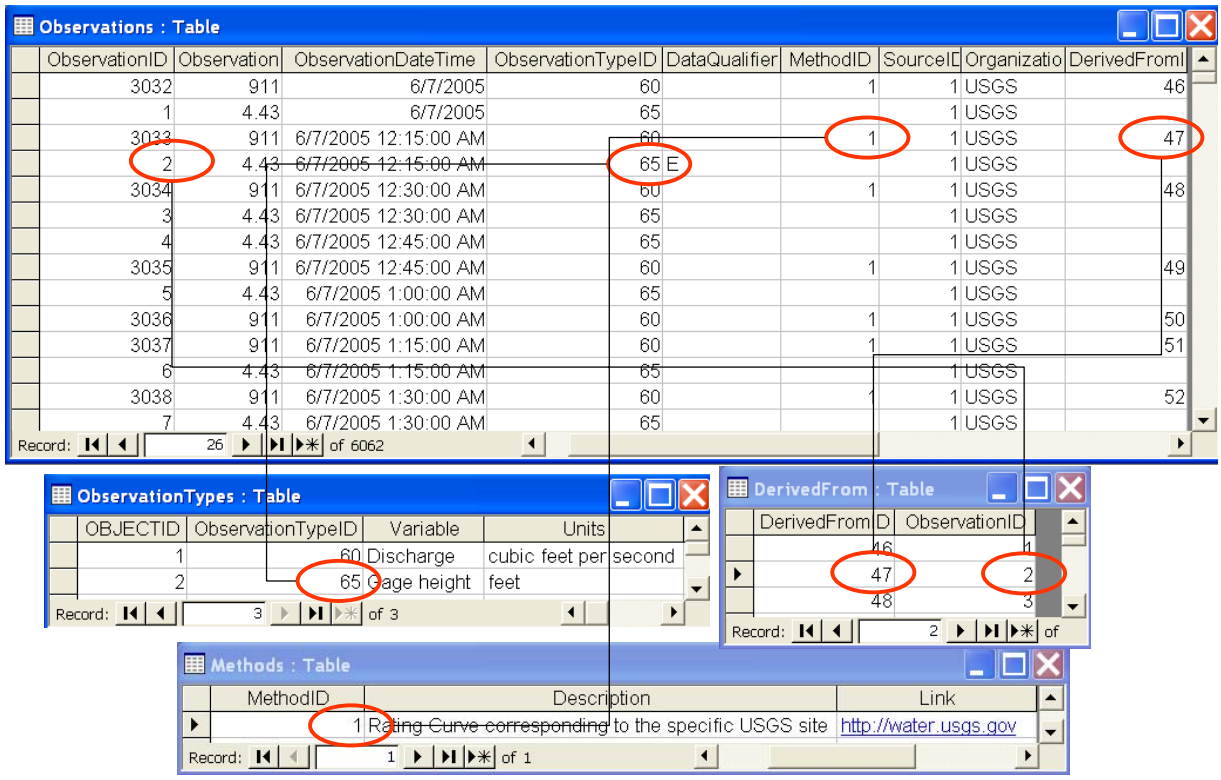


Figure A.1. Excerpts from tables illustrating the population of the data model with streamflow stage and discharge data.

Note that stage in feet and discharge in cubic feet per second are both in the same data table but with different observation types that reference the variable, units and other quantities associated with these observations. The link between ObservationTypeID in the Observations table and ObservationTypes table is shown. In this example, discharge measurements are presumed to be derived from stage measurements through a rating curve. The MethodID associated with each Discharge record references into a method table that describes this and provides a URL that should contain metadata details for this method. The DerivedFromID in the Observations table references into the DerivedFrom table that references back to the corresponding stage in the observations table from which the discharge was derived.

Water Chemistry from a Profile in a Lake

Reservoir profile measurements provide an example of observations that should logically be grouped and observations that have an offset in relationship to the location of the sampling station. These measurements may be made simultaneously (by multiple instruments in the water column) or over a short time period (one instrument that is lowered from top to bottom). The following shows an example of how these data would be stored in the proposed database structure.

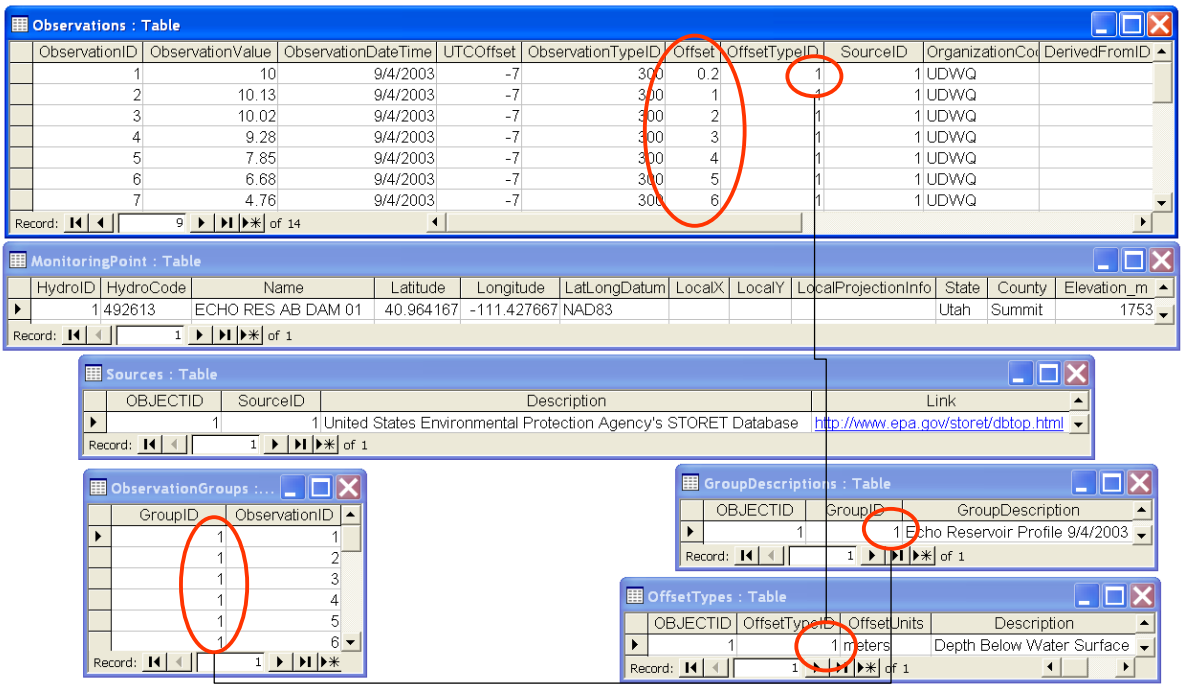


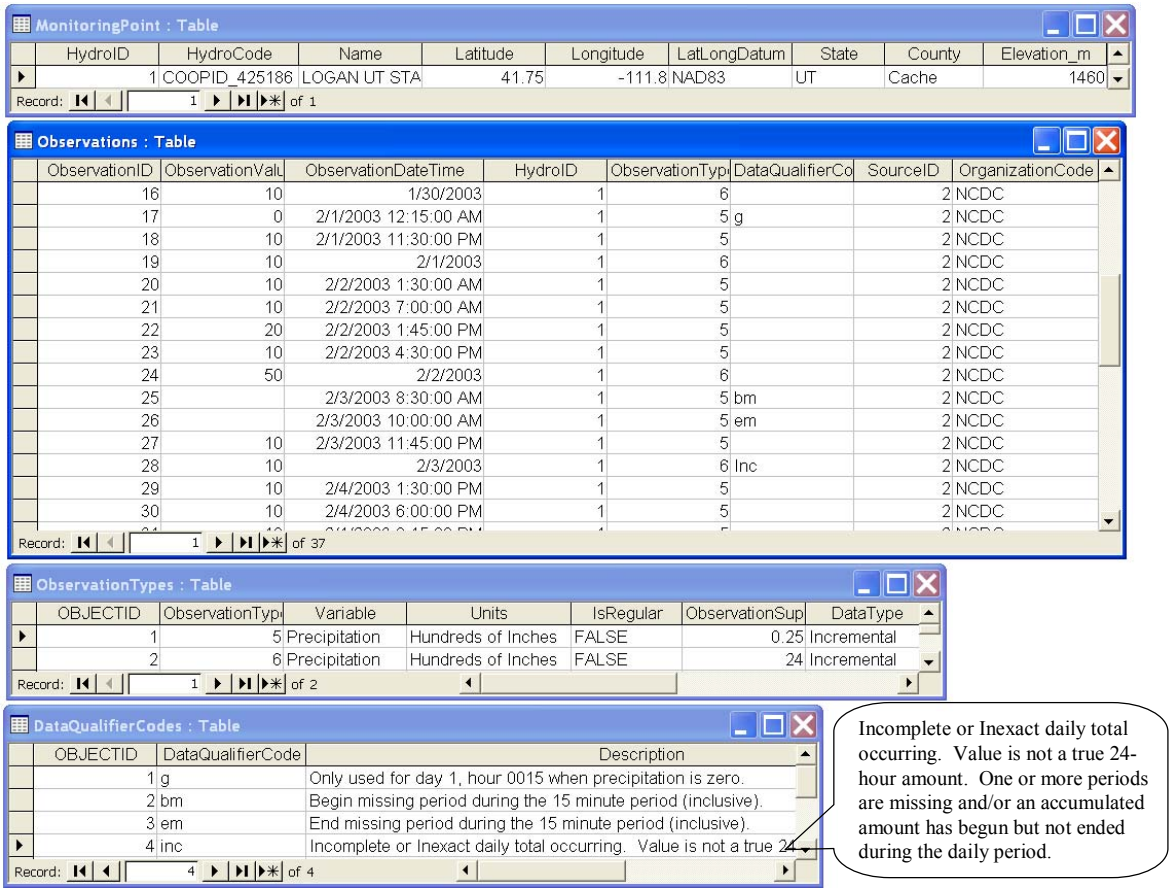
Figure A.2. Excerpts from tables illustrating the population of the data model with Water Chemistry data.

This example illustrates the use of the OffsetTypes table and Offset attribute to quantify the depth associated with each measurement. This example also illustrates the use of the ObservationGroups table and GroupDescriptions table to group logically related measurements. The MonitoringPoint table includes HydroID and shape information (not shown) that locates each observation geographically within a GIS, but also includes Latitude and Longitude and LocalX and LocalY coordinates to provide location information independent of the GIS system. The Sources table indicates the source of this data from the EPA STORET database with URL given.

NCDC Precipitation Data

Figure A.3 illustrates the representation of NCDC 15 min precipitation data by the Data Model. The data files include 15 min observations as well as daily totals. Separate observation types are used for the 15 min or daily total values. This data is reported at irregular intervals and only for time periods for which precipitation is non zero. This is accommodated by setting the IsRegular attribute associated with the observation type to False and specifying the ObservationSupport

value as 0.25 hr or 24 hr. The DataType of 'incremental' is used to indicate that these are incremental values defined over the ObservationSupport interval. Data qualifier codes indicate periods where the data is missing. This is necessary because of the convention that zero precipitation periods are not reported. A data qualifier code is also used to flag days where the precipitation total is incomplete due to the record being missing during part of the day.



MonitoringPoint : Table

HydroID	HydroCode	Name	Latitude	Longitude	LatLongDatum	State	County	Elevation_m
1	COOPID_425186	LOGAN UT STA	41.75	-111.8	NAD83	UT	Cache	1460

Record: 1 of 1

Observations : Table

ObservationID	ObservationValu	ObservationDateTime	HydroID	ObservationTypi	DataQualifierCo	SourceID	OrganizationCode
16	10	1/30/2003	1	6		2	NCDC
17	0	2/1/2003 12:15:00 AM	1	5	g	2	NCDC
18	10	2/1/2003 11:30:00 PM	1	5		2	NCDC
19	10	2/1/2003	1	6		2	NCDC
20	10	2/2/2003 1:30:00 AM	1	5		2	NCDC
21	10	2/2/2003 7:00:00 AM	1	5		2	NCDC
22	20	2/2/2003 1:45:00 PM	1	5		2	NCDC
23	10	2/2/2003 4:30:00 PM	1	5		2	NCDC
24	50	2/2/2003	1	6		2	NCDC
25		2/3/2003 8:30:00 AM	1	5	bm	2	NCDC
26		2/3/2003 10:00:00 AM	1	5	em	2	NCDC
27	10	2/3/2003 11:45:00 PM	1	5		2	NCDC
28	10	2/3/2003	1	6	Inc	2	NCDC
29	10	2/4/2003 1:30:00 PM	1	5		2	NCDC
30	10	2/4/2003 6:00:00 PM	1	5		2	NCDC

Record: 1 of 37

ObservationTypes : Table

OBJECTID	ObservationTypi	Variable	Units	IsRegular	ObservationSup	DataType
1	5	Precipitation	Hundreds of Inches	FALSE	0.25	Incremental
2	6	Precipitation	Hundreds of Inches	FALSE	24	Incremental

Record: 1 of 2

DataQualifierCodes : Table

OBJECTID	DataQualifierCode	Description
1	g	Only used for day 1, hour 0015 when precipitation is zero.
2	bm	Begin missing period during the 15 minute period (inclusive).
3	em	End missing period during the 15 minute period (inclusive).
4	inc	Incomplete or Inexact daily total occurring. Value is not a true 24

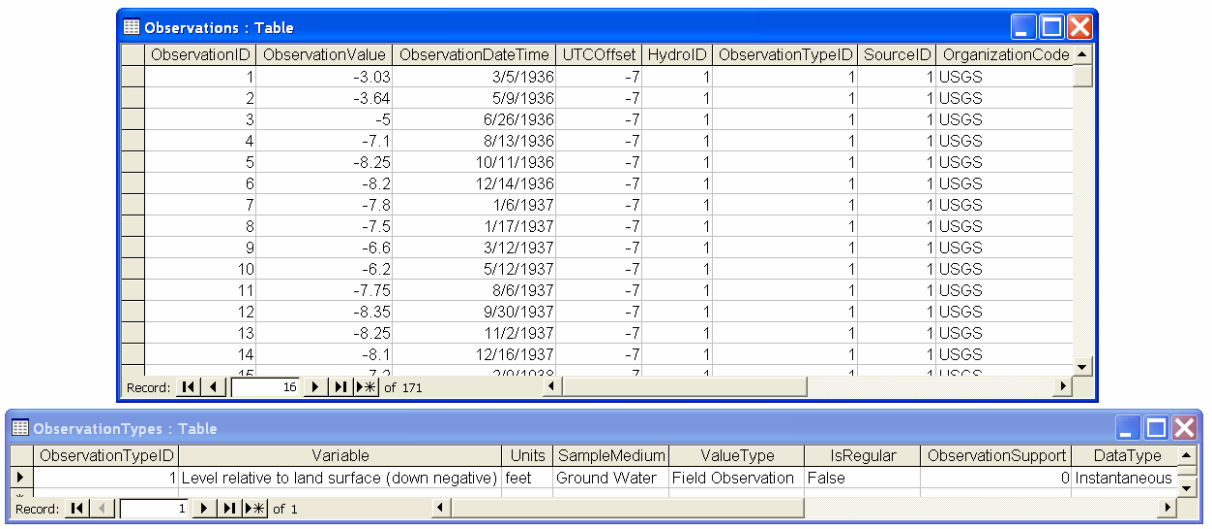
Record: 4 of 4

Incomplete or Inexact daily total occurring. Value is not a true 24-hour amount. One or more periods are missing and/or an accumulated amount has begun but not ended during the daily period.

Figure A.3. Excerpts from tables illustrating the population of the data model with NCDC Precipitation Data.

Groundwater Level

The following is an example of how groundwater level data can be stored in the proposed database structure.



Observations : Table

ObservationID	ObservationValue	ObservationDateTime	UTCOffset	HydroID	ObservationTypeID	SourceID	OrganizationCode
1	-3.03	3/5/1936	-7	1	1	1	USGS
2	-3.64	5/9/1936	-7	1	1	1	USGS
3	-5	6/26/1936	-7	1	1	1	USGS
4	-7.1	8/13/1936	-7	1	1	1	USGS
5	-8.25	10/11/1936	-7	1	1	1	USGS
6	-8.2	12/14/1936	-7	1	1	1	USGS
7	-7.8	1/6/1937	-7	1	1	1	USGS
8	-7.5	1/17/1937	-7	1	1	1	USGS
9	-6.6	3/12/1937	-7	1	1	1	USGS
10	-6.2	5/12/1937	-7	1	1	1	USGS
11	-7.75	8/6/1937	-7	1	1	1	USGS
12	-8.35	9/30/1937	-7	1	1	1	USGS
13	-8.25	11/2/1937	-7	1	1	1	USGS
14	-8.1	12/16/1937	-7	1	1	1	USGS

Record: 16 of 171

ObservationTypes : Table

ObservationTypeID	Variable	Units	SampleMedium	ValueType	IsRegular	ObservationSupport	DataType
1	Level relative to land surface (down negative)	feet	Ground Water	Field Observation	False	0	Instantaneous

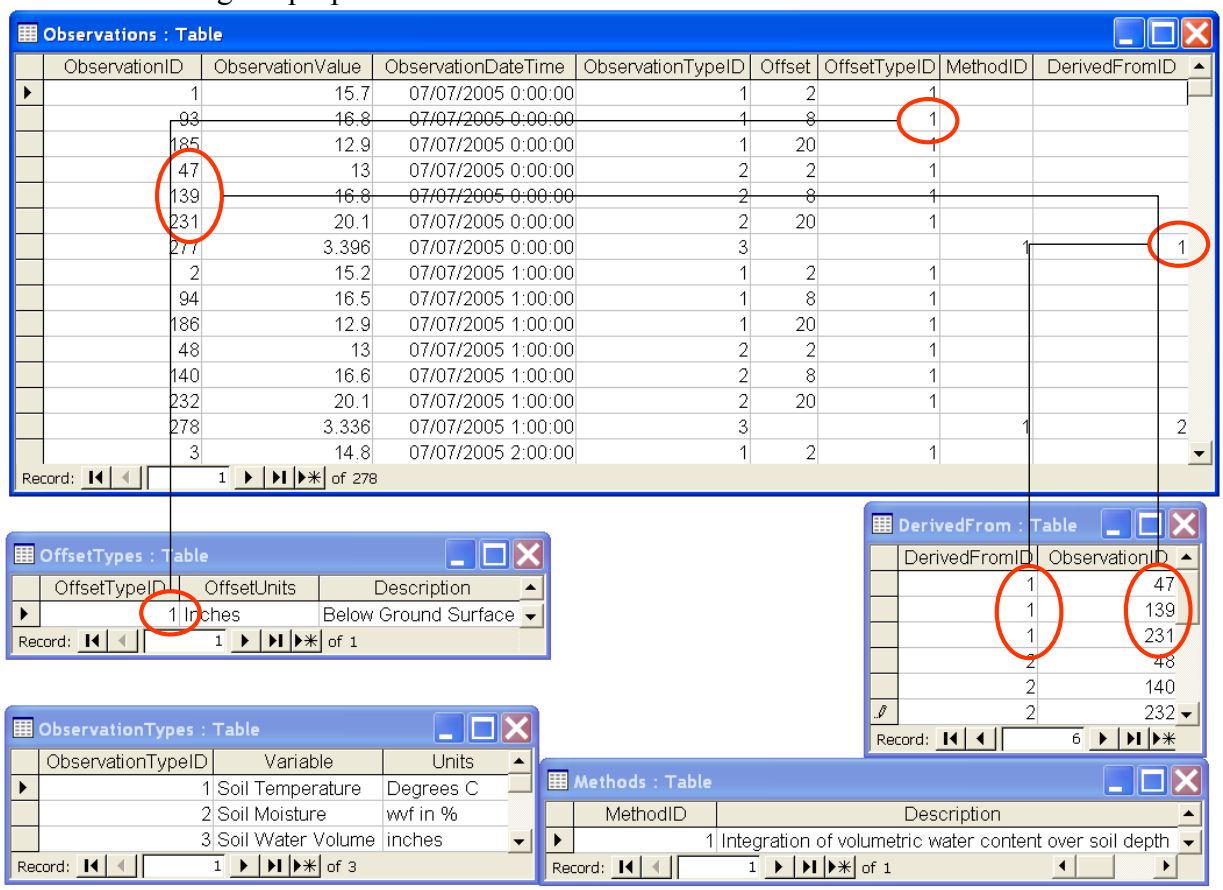
Record: 1 of 1

Figure A.4. Excerpts from tables illustrating the population of the data model with irregularly sampled groundwater data.

In this groundwater level example observations are depth relative to the ground surface reported as negative values.

Soil Moisture Sampled from a Depth

Soil moisture and soil temperature are examples of quantities that may be measured over a range of depths at a sampling location. The following (Figure A.6) is an example of how these data can be stored using the proposed database structure.



Observations : Table

ObservationID	ObservationValue	ObservationDateTime	ObservationTypeID	Offset	OffsetTypeID	MethodID	DerivedFromID
1	15.7	07/07/2005 0:00:00	1	2	1		
93	16.8	07/07/2005 0:00:00	1	8	1		
185	12.9	07/07/2005 0:00:00	1	20	1		
47	13	07/07/2005 0:00:00	2	2	1		
139	16.8	07/07/2005 0:00:00	2	8	1		
231	20.1	07/07/2005 0:00:00	2	20	1		
277	3.396	07/07/2005 0:00:00	3			1	1
2	15.2	07/07/2005 1:00:00	1	2	1		
94	16.5	07/07/2005 1:00:00	1	8	1		
186	12.9	07/07/2005 1:00:00	1	20	1		
48	13	07/07/2005 1:00:00	2	2	1		
140	16.6	07/07/2005 1:00:00	2	8	1		
232	20.1	07/07/2005 1:00:00	2	20	1		
278	3.336	07/07/2005 1:00:00	3			1	2
3	14.8	07/07/2005 2:00:00	1	2	1		

Record: 1 of 278

OffsetTypes : Table

OffsetTypeID	OffsetUnits	Description
1	Inches	Below Ground Surface

Record: 1 of 1

DerivedFrom : Table

DerivedFromID	ObservationID
1	47
1	139
1	231
2	48
2	140
2	232

Record: 6

ObservationTypes : Table

ObservationTypeID	Variable	Units
1	Soil Temperature	Degrees C
2	Soil Moisture	wvf in %
3	Soil Water Volume	inches

Record: 1 of 3

Methods : Table

MethodID	Description
1	Integration of volumetric water content over soil depth

Record: 1 of 1

Figure A.5. Excerpts from tables illustrating the population of the data model with soil moisture and temperature data collected over a profile into the soil.

In this example at each DateTime there are 3 measurements of soil moisture at depths (2, 8 and 20 inches) and 3 measurements of soil temperature at these same depths. The OffsetTypes table indicates that these measurements refer to depth below the ground. There is a single derived soil water volume obtained by integrating soil moisture over the soil profile. The methods table describes the method and the DerivedFrom table gives the groups of three soil moisture measurements that were used in deriving each soil water volume value, illustrating how this model works when groups of variables are used in obtaining a derived quantity.