CUAHSI

universities allied for water research

A NOTE ON HIS CENTRAL 1.2 CATALOG UPDATE WITH USGS NWIS INFORMATION, AND SEMANTIC VARIABLE TAGGING

April 25, 2010

by:

Ilya Zaslavsky,
Tom Whitenack,
David Valentine
San Diego Supercomputer Center
University of California, San Diego

Distribution

Funding

This document was developed as part of the Consortium of Universities for the Advancement of Hydrologic Science, Inc. (CUAHSI) Hydrologic Information System project, under NSF Grant No. EAR-0413265. We thank for their input and insights our colleagues from USGS: David Briar, Jon Scott and Nate Booth.

TABLE OF CONTENTS

Introduction

HIS Central is the web application which provides an interface for adding and managing registered water data services and the HIS central metadata catalog. The metadata catalog is updated from many sources. The two groups of sources include: ODM-based sources that implement a standard suite of WaterOneFlow services, and sources where WaterOneFlow services have been developed to expose the content of a remote data repository, typically as a wrapper of the external web site or services. In the former case, catalog harvesting is done automatically and regularly, as described in the main HISCentral document. In the latter case, HISCentral harvests the catalog periodically over an independently updated replica of the remote catalog. The methods of creating such replicas differ by agency. In the case of EPA, the entire content of STORET is provided to SDSC as Oracle database dump. Tables in this dump are then imported into an MS SQL Server instance and transformed to comply with an ODM-like catalog schema which is then harvested by the HIS Central application. The procedure is described in detail in the main HIS Central 1.2 document. In the case of USGS, selected catalog tables are exported for use by SDSC, and converted into an ODM-like catalog schema in MS SQL Server, using a collection of specially designed views. This process is described in the first section of this document.

Typically, sources of the second type also have extensive and rich semantic content, which requires that the variables (parameters) are associated with concepts in a CUAHSI-maintained ontology. The second part of this document analyzes this semantic tagging procedure and results, as produced in March-April of 2010, as a guide for further work on semantic annotation.

The source data for the analysis are provided by the NWIS catalog dump from April 15, 2010. The ontology tagging of SRS-crosswalked NWIS parameter codes is developed at the CUAHSI Central Office. Jon Scott and Nate Booth contributed to the interpretation of the findings. We are grateful for their help.

NWIS catalog import, and its harvesting into HIS Central

The NWIS catalog dump consists of 6 tables. Their names and record numbers are listed below.

Table/file	Count	Description
USGS_sitefile	1545778	All sites. Shared between services
USGS_Param_list	18615	
USGS_POR_List	1072497	Real Time station flag, plus generalized period of record listing (count, begin/end) for groundwater, peak flow , non-specific qw (all Intermittent Irregular data).
usgs_qw_por_List	14855675	Detailed Period of Record for Intermittent Irregular data: Site-parameter-begin-end-count.
USGS_DV_DD_POR_list	68024	Detailed Period of Record for Daily Values aggregates: site-instrument-parameter-statistic-begin-end-count. There can be more than one instrument providing the same parameter at a station)
USGS_UV_DD_List	28928	Detailed Period of Record for Real Time: site-instrument-parameter . Real time has a limited time period for availability,

		so dates can be created in a view. There can be more than one instrument providing the same parameter at a station.

Once the dump files are downloaded from the USGS ftp site, they are ingested in an SQL Server instance. Date fields in the USGS DB use 00 to indicate precision, so dates are ingested as text, and update commands are run to create date time fields. For example this updates groundwater datetime fields:

```
Update usgs_por_list
-- This creates datetime fields
-- and addresses issues where month and year data are reperesented
--  XXXX-00-00 year; start 01-01 end 12-31
-- XXXX-XX-00 month start 01 end 28 (simplified calculation)
  Set gw_BeginDateTime = ( case
   when( gw_begin_date < '1850') then null
      when( gw_begin_date like '0000-00-00') then null
      when( gw_begin_date like '%-00-00') then  cast
(substring(gw_begin_date, 0,5) + '-01-01' as date)
     when( gw_begin_date like '%-00') then  cast (substring(gw_begin_date,
0,8) + '-01'  as date)
      else gw_begin_date
      end ),
gw_EndDateTime = (case
     when( gw_end_date < '1850') then null
      when( gw_end_date like '0000-00-00') then null
      when( gw_end_date like '%-12-31') then  cast (substring(gw_end_date,
0,5) + '-01-01' as date)
     when( gw_end_date like '%-28') then  cast (substring(gw_end_date, 0,8) +
'-01'  as date)
      else gw_end_date
      end)
-- If count > 0 there is data
  where gw_count_nu > 0
```

Then a collection of views is used to 1) present the tables as ODM-like tables, 2) create additional tables (e.g. the variables table is created from several period of records tables, unioning variables as expressed in qw, gw, uv and dv POR tables). The key views are listed in the appendix.

Specific issues with the USGS catalog dump, as discussed with Jon Scott and Nate Booth, include: a) it contains station codes but not station identifiers (including organization). Sometimes we need both to unambiguously identify a station; b) it doesn't contain medium information.

Once the ODM-like replica of the NWIS catalog is created, it is harvested into the HIS Central metadata catalog as described in the HIS Central 1.2 documentation.

Analysis of NWIS parameter mappings

To assess the success of the ontology tagging we need to establish what we are trying to compare on the catalog side, and on the ontology side, respectively. The catalog update process is presented above, and it produced 4853

unique parameter names that have associated time series. These parameter names were distilled from period of record tables for each of the 4 NWIS subsystems, namely qw, dv, uv and gw (*POR* tables) obtained as part of NWIS catalog dump. Therefore, this section focuses on distilling a comparable set of codes from SRS tables, so that these parameters can be matched with the parameters in the catalog dump.

PRELIMINARY NOTES

A conversation with Jon Scott (the developer of the SRS-based coordination on parameter codes between USGS and EPA nomenclatures, http://qwwebservices.usgs.gov/public_srsnames.html) and Nate Booth (the developer of the USGS water quality service, http://qwwebservices.usgs.gov/), revealed the following expected differences between parameter codes in the catalog dump, and in SRS.

Potential reasons for parameter codes to appear in the catalog but not mapped to SRS:

- Observation metadata (e.g. weather) and various characteristics of measurement context are not in the SRS. Historically they have been added as parameter codes, but really should be treated as attributes of measurements. So they appear in the catalog dump but not in the SRS.

- Some variables in the EPA list have unconventional semantics, and were excluded from the SRS – but may be in the catalog. There should be just a few of those.

- Surrogate parameters (those added in the lab, not field measurements, describing performance of analytical processes and usually measured in %% recovery) are not in SRS, but included in the catalog (there may be several hundred of those)

Potential reasons for parameter codes to appear in SRS but not in the catalog:

- Not all parameter codes listed in the SRS, have associated data. This will account for the largest discrepancy between the catalog and the SRS lists. Nate Booth provided a list of SRS parameters that have associated data (hereafter referred to as the USGS list).

ANALYSIS OF AVAILABLE SRS TAGGING

This section presents a step-by-step analysis of the parameter codes as provided in SRS and tagged at the CUAHSI Central office.

1. There are 9178 codes total in the SRS crosswalk table on the web, which provides mappings between SRS concepts and NWIS parameter codes (source: http://qwwebservices.usgs.gov/public_srsnames.html)

   Overall the SRS list of concepts that have associated data, contains 1139 concepts (source: http://qwwebservices.usgs.gov/html/item_select.html?caller=characteristicName&previousSelections=)

2. The list of SRS-crosswalked NWIS parameters tagged at the CUAHSI Central Office contains 8408 codes. Hereafter, we refer to this list as CUAHSI Ontology Table (COT). Of those parameter codes, 3567 received an ontology tag (source: COT from February 26, 2010).

Issue 1. The list used in COT, doesn't appear to be a subset of the SRS-crosswalked list. There are 54 codes that appear in COT and don't appear in the SRS crosswalk table. It is not clear where they came from.

Issue 2. Matching the COT list of 8408 with the SRS crosswalk list of 9178, produced 8273 codes that appear in both lists. The additional difference comes from the fact that 81 of parameter codes in the COT list have duplicate records.

Issue 3. We have to work with unique NWIS codes that exist in the SRS crosswalk, hence our starting number is 8273.

I hope these differences do not affect the quality of tagging; they just lower the number of parameter codes that we can use for estimating the success of tagging.

3. Of the 9178 parameter coded in the SRS crosswalk list, 4339 parameter codes have associated data (source: list of NWIS codes with data provided by USGS, referred to as the USGS list). The remaining 4838 NWIS parameter codes don't have associated data and therefore are not expected to have matches in the NWIS catalog dump, or are information parameters (project codes, set codes, sample numbers) and don't expect to be searchable in CUAHSI HIS.

4. Among the 4339 parameter codes that have data, we identify 3 groups:

    - Have data, and tagged in COT: 2106 parameters (this is essentially our basis for comparison when evaluating catalog tagging)

      Browsing the parameter codes reveals that these are mostly measurements on water and suspended sediment. However, parameters such as air temperature, wind direction, barometric pressure, radiation, reservoir storage in acre feet, odor, stream width, biomass (periphyton, dry or ash weight), ice cover, dead fish, odor etc. severity codes, depth of flow in pipe, temperature in soil, depth of lake; ice thickness, snow depth, sieve diameter and fall diameter in sediments; phytoplankton, specific gravity, depth to water level – are also present. Given this variety, it is not clear what criteria were used to select parameters for tagging.

    - Have data, but not tagged in COT: 2046

      Browsing these parameters revealed that they include a large number of variables measured in other mediums: bed sediment, soil, rock, biota (tissue, whole organism). However, many measurements in water and suspended sediment are also present. Measurement of atmospheric deposition (dry, wet and bulk) are also mostly in this group. In addition, this group has a number of parameters reflecting counting errors and uncertainties, but also parameters like altitude. For many variables, it is not clear why they are in this group and not in the first one. I believe this

group of parameter codes is the single most important area where further tagging should happen.

- Have data, but not included in the COT list of 8408: 188

  I didn't find any special pattern that can be used to explain their exclusion from the tagging list. Parameters here include measurements in water, suspended solids, bed sediments, suspended sediments, biota, soils. On many of these parameters USGS and EPA are in agreement. These parameters should be considered for future tagging.

5. Among the 4838 parameters that don't have associated data, we identify the same 3 groups:

   - Don't have data, but nevertheless tagged in COT: 1385

     These codes are not expected to have matches in the catalog dump, so it is not clear why they were tagged. From a cursory browsing it appears that some of these parameters (measured mostly in water and suspended sediment), have verbatim parameters in more common units already mapped. In this group, we often have the same parameters but in less commonly used units (e.g. Flow, gallons per batch; picocuries per liter; barometric pressure in many different units; dissolved calcium tons per day). Many radioactive metals and organics that are found in very small concentrations (e.g. pharmaceuticals) are also here.

   - Don't have data, and not tagged in COT: 2736

     Many of the same patterns as above, including counting errors, but also many in all mediums

   - Don't have data, and not included in the COT list of 8408: 717

     No discernable pattern here either, similar to the third category in the previous group.

ANALYSIS OF MATCHING TAGGED SRS PARAMETERS WITH PARAMETERS IN THE CATALOG DUMP

Our next step is to match the 2106 unique and valid NWIS codes associated with SRS concepts (i.e. the codes are present in the SRS list, have data, and are tagged in COT) against the parameter codes present in the NWIS catalog dump.

In the v_iz_unique_params_from_dump, a view that contains a list of unique parameter codes that have associated periods of record in qw, uv or dv subsystems of NWIS (7218 records), we observe (using the same groupings as before):

- 2103 codes are matched with SRS, have data, and have been tagged in COT. Therefore, the tagging accuracy is nearly 100%. The three parameter codes that have been tagged in COT, have data in the SRS, but were not found in the catalog, are: 1) # 29797, Bicarbonate, water, unfiltered, Gran titration, field, milligrams per liter, b) # 29798, Carbonate, water, unfiltered, Gran titration, field, milligrams per liter, c) # 00909, Hardness, water, unfiltered, measured, milligrams per liter as calcium carbonate.
- There are 2046 codes which are matched between the catalog dump and the SRS, have data , but have not been tagged in COT. This is a big group worth exploring, as potential extension of the tagging work. The measurements are mostly in other mediums though water and suspended sediment are also present here.
- 188 codes are matched with SRS, have data, but are not included in the COT list. No medium-dependent patterns were identified for the content of this list. Many PCBs and isomers are here.
- 25 parameter codes have data in the catalog dump and are matched with SRS, but don't have data according to the list provided by USGS, and mapped in COT. These parameters include: 4 barometric pressure codes; precipitation mm; flow rate mil gallons/day; tide stage; solar radiation, reservoir storage, a number of various small organics (pharmaceuticals). It could be an issue of units, and the "uncorrected" status present in some parameter description – which excluded them from the "SRS-crosswalked with data" group.
- 8 codes are matched with SRS, don't have data according to the USGS list, and are not mapped in COT. Again, these are small organics (pharmaceuticals, e.g. testosterone, androsterone, coprostanol in water), also distance to stream bottom, suspended sediment concentration.
- 18 codes are matched with SRS, don't have data according to the USGS list, and are not included in the COT list. This group includes various metals measured in bed sediment (aluminum, iron, lead, lithium, manganese, molybdenum, nickel, silver, strontium, tin, titanium, uranium, zinc), and a few additional characteristics measured in water and soil.

This covers 4388 parameter codes in the NWIS catalog dump. Therefore, we have 2830 codes (2830=7218-4388) which exist in the data dump, but don't show up in the SRS crosswalk. Here is the composition of these codes by NWIS information categories is:

- information: 224 (land use, sample weight, exposure, lot number, set number, water use, possible contamination source, transit rate, location, azimuth from outlet, type of well, conversion factor, agency analyzing sample, weather (WMO code), battery voltage, instrument orientation, age of organism, signal to noise,
- biological: 29 (invertebrates, male and female counts, productivity, e-coli, coliphage, bacteria, etc.)
- major inorganics: 2 (organic carbon/organic nitrogen ratio; nitrogen plus carbon monoxide)
- minor and trace inorganics: 16 (arsenic, mercury, delta helium, arsenite, arsenate, cyanide, iron (reactive))
- nutrients: 6 (hydrolizable phosphorus – several measures, albumimoid nitrogen)
- organics: 315 (many surrogates here; but not only them; various organics incl. penicillin, amoxicillin, metformin, ampicillin, menthol, tetracycline, DDT)

- other: 14 (“Water Science Center Special”)
- physical property: 154 (bulk density, radiation, evaporation temperature, location, discharge, surface area; precipitation, wet bulb t, % organic matter, resistivity, illumination, time of travel; oxygen demand; latent heat flux; wind gust direction, etc.)
- radiochemical: 8 (apparent age, isotopes of radium, strontium ration, uranium ration,
- sediment: 26 (solids, %% smaller than various thresholds)
- codes not found in the list of available parameter codes: 2036.

Many parameters in this group reflect the exclusion criteria used in creating the SRS crosswalk list, as mentioned by Jon Scott (see the Preliminary Notes section). Jon Scott further clarifies the mismatch between the NWIS catalog dump and the SRS-matched parameter codes as follows (personal communication, 4/27/10):

> To assist you in understanding parameter codes for which USGS has data (from Dave Briar's inventory) and their lack of relationship to mappings of parameters on the QW web services, I produced the following inventory of parameters that have data but are not mapped. I aggregated the unmapped results into constituent groups and counted the parameter codes. For the most populous groups, I have added some explanations to the side. You'll recall our introductory remarks about intentionally omitting the taxonomic organism counts due to lack of expertise and decades old taxonomic nomenclature.

GROUP	NUMBER OF PARAMETERS	DESCRIPTION
POP	1981	taxonomic identification, benthics, algae, etc.
INF	278	metadata, such as analytical set number, surrogate recovery
OOT	83	Other organics, lacking SRS name, esp. pharamceuticals
PHY	82	Physical observations during site visits
OPE	79	Pesticides, no matching SRS name established
OTH	27	Poorly defined parameters, not made publicly available by USGS
BIO	26	
SED	22	
MBI	16	
ISO	7	
NUT	7	
IMM	5	
IMN	4	
RAD	4	
INN	1	

> Note that according to Jon Scott, the above data reflect an inventory of USGS parameter usage of several years ago, with 7012 unique parameter codes. This explains the numerical difference between mismatches reported by Jon, and calculated from the recent catalog dump. The explanatory patterns, however, remain the same.

CONCLUSION

The accuracy of matching is nearly 100% (2103 out of 2106). The completeness of matching (2103 out of 7218) is 29%. Further work is required to tag parameter codes in other mediums, to increase the completeness of mapping

parameters in the catalog. However, the tagging work should focus on those SRS-matched USGS parameter codes that have data (4340 parameters, 2106 of which are already tagged to ontology concepts). One option for doing this is extending tagging to variables whose names are verbatim to already tagged variables, except for units or medium. Also, future tagging work should address those parameters in the catalog that are not listed in the SRS crosswalk, except for those parameters that are not likely to be requested via HIS searches.

Matching variables in the catalog with variables in ontology by variable names rather than by parameter codes is an alternate option that can be complementary to the procedure described above. Following this method, variable names in the catalog are matched with variable names from the CUAHSI controlled vocabulary, which, in turn, are tagged to ontology terms. While this approach does not take into account medium or units of measure, this may not be needed for initial search operations as long as the medium and units are returned in service calls.

APPENDIX A. VIEWS USED TO CONSTRUCT ODM TABLES FROM THE USGS CATALOG DUMP.

For each of the four services, a set of three views is created: sites, series, and variables. These views each have some customized information. For the daily values view, a variable with the statistical code is created, and time support is different for each source.

Source	isRegular	TimeSupport	TimeUnits	Network
UV (Real Time)	T	15	M	NWISUV
DV (Daily Values)	T	1	D	NWISDV
QW (IID)	F	0	H	NWISIID
GW (groundwater levels)	F	0	H	NWISGW

Example Daily Values Statistics

stat_cd	USGSDescription	DataType (CUAHSI Name)	Example Code
00001	Maximum	Maximum	00003/DataType=Maximum
00002	Minimum	Minimum	00003/DataType=Minimum
00003	Mean	Average	00003/DataType=Average
00004	AM	NULL	00010/statistic=00004
00006	Sum	Cumulative	00036/DataType=Cumulative
00008	Median	Median	00011/DataType=Median
00009	STD	NULL	00010/statistic=00009
00011	Instantaneous	Instantaneous	00010/DataType=Instantaneous

```
-- Sites
SELECT     dbo.USGS_sitefile.site_id AS SiteID, dbo.USGS_sitefile.site_no AS
SiteCode, dbo.USGS_sitefile.station_nm AS SiteName,
dbo.USGS_sitefile.dec_lat_va AS Latitude,
```

```
                      dbo.USGS_sitefile.dec_long_va AS Longitude,
dbo.SpatialReferences.SpatialReferenceID AS LatLongDatumID,
dbo.USGS_sitefile.alt_va AS Elevation_m,
                      dbo.USGS_sitefile.alt_datum_cd AS VerticalDatum, NULL
AS PosAccuracy_m, dbo.v_fips_countyCodes.County,
dbo.v_fips_countyCodes.State, NULL AS Comments,
                      dbo.USGS_sitefile.tz_cd AS odws_timeZoneName,
dbo.USGS_sitefile.local_time_fg AS odws_UsesDaylightSavingsTime,
                      dbo.USGS_sitefile.state_cd AS usgs_State_fipCode,
dbo.USGS_sitefile.county_cd AS Usgs_County_fipsCode,
dbo.USGS_sitefile.agency_cd AS usgs_agency,
                      dbo.usgs_station_type_cd.usgs_station_type, 'NWISIID'
AS SiteVocabulary, 'NWISIID' AS VariableVocabulary, 'NWISIID:' +
dbo.USGS_sitefile.site_no AS hs_SiteCode,
                       dbo.USGS_sitefile.huc_cd AS usgs_huc, CASE len(HUC_CD)
WHEN 0 THEN NULL ELSE dbo.USGS_sitefile.huc_cd + CASE
len(rtrim(ltrim(basin_cd)))
                      WHEN 0 THEN '0000' WHEN 2 THEN Basin_cd + '00' WHEN 4
THEN Basin_cd ELSE '0000' END END AS HUCNUMERIC
FROM         dbo.USGS_sitefile INNER JOIN
                      dbo.qw_SiteCodes_used AS sitesUsed ON
dbo.USGS_sitefile.site_id = sitesUsed.siteID LEFT OUTER JOIN
                      dbo.usgs_station_type_cd ON
dbo.USGS_sitefile.station_type_cd = dbo.usgs_station_type_cd.station_type_cd
LEFT OUTER JOIN
                      dbo.v_fips_countyCodes ON dbo.USGS_sitefile.county_cd =
dbo.v_fips_countyCodes.FIPS_county_cd AND
                      dbo.USGS_sitefile.state_cd =
dbo.v_fips_countyCodes.FIPS_st_cd LEFT OUTER JOIN
                      dbo.SpatialReferences ON
dbo.USGS_sitefile.dec_coord_datum_cd = dbo.SpatialReferences.Name
WHERE     (NOT (dbo.USGS_sitefile.dec_lat_va = 0)) AND (NOT
(dbo.USGS_sitefile.dec_long_va = 0))
```

```
-- Variables
-- if we had a table with odm properties (medium, valuetype) we could
populate those fields
SELECT     TOP (100) PERCENT dbo.variables.VariableID,
usedVariables.VariableCode, CASE WHEN dbo.NwisVariableName.variableName IS
NULL
                      THEN variables.variablename ELSE
dbo.NwisVariableName.variableName END AS VariableName, 'Unknown' AS
SampleMedium, 'Unknown' AS ValueType,
                      'false' AS IsRegular, '0' AS TimeSupport, '104' AS
TimeUnitsID, 'Sporadic' AS DataType, dbo.variables.GeneralCategory,
dbo.variables.NoDataValue,
                      dbo.variables.networkId, dbo.variables.VariableUnitsID,
dbo.variables.mediumCode, dbo.NwisVariableName.variableDescription AS
usgs_parsedVariableDescription,
```

```
                      dbo.NwisVariableName.variableOrigText AS
usgs_variableDescription, 'NWISIID' AS VariableVocabulary, 'NWIS:' +
dbo.variables.VariableCode AS hs_variableCode,
                      'NWISIID:' + dbo.variables.VariableCode AS
hs_newVariableCode, 'd' AS TimeUnitsAbbreviation, 'day' AS TimeUnitsName,
                      dbo.Units.UnitAbbreviation AS
VariableUnitsAbbreviation, dbo.Units.Units AS VariableUnitsName,
dbo.Units.UnitType AS VariableUnitsType
FROM         dbo.variables INNER JOIN
                          (SELECT     PARAMETER_CD AS VariableCode,
COUNT(PARAMETER_CD) AS Count
                            FROM          dbo.usgs_qw_por
                            GROUP BY PARAMETER_CD) AS usedVariables ON
dbo.variables.VariableCode = usedVariables.VariableCode LEFT OUTER JOIN
                      dbo.Units ON dbo.variables.VariableUnitsID =
dbo.Units.UnitID LEFT OUTER JOIN
                      dbo.NwisVariableName ON usedVariables.VariableCode =
dbo.NwisVariableName.variableCode
WHERE     (dbo.variables.VariableID IS NOT NULL)
ORDER BY usedVariables.VariableCode
```

```
-- Series
SELECT     dbo.USGS_sitefile.site_id AS siteID, dbo.USGS_sitefile.site_no AS
SiteCode, dbo.USGS_sitefile.station_nm AS siteName, dbo.Units.Units AS
VariableUnitsName,
                      dbo.Units.UnitID AS VariableUnitsID, Units_1.Units AS
TimeUnitsNmae, NULL AS MethodID, CONVERT(varchar(255), NULL) AS MethodName, 2
AS SourceID,
                      'USGS' AS Organization, 'USGS NWIS IID' AS
SourceDescription, NULL AS QualityControlLevelID, sitePOR.BEGINDDATE AS
BeginDateTime,
                      sitePOR.ENDDATE AS EndDateTime, sitePOR.VALUECOUNT,
'NWISIID' AS VariableVocabulary, 'NWISIID' AS SiteVocabulary, 'Instantaneous'
AS DataType,
                      dbo.USGS_sitefile.agency_cd AS usgs_agency,
variables.hs_newVariableCode AS ws_variablecode,
variables.TimeUnitsAbbreviation, variables.TimeUnitsName,
                      variables.SampleMedium, variables.ValueType,
variables.IsRegular, variables.TimeSupport, variables.TimeUnitsID,
variables.GeneralCategory,
                      variables.VariableName, variables.VariableCode, 'NWIS:'
+ variables.VariableCode AS hs_variableCode, 'NWIS:' +
dbo.USGS_sitefile.site_no AS hs_siteCode,
                      'NWISIID:' + dbo.USGS_sitefile.site_no AS ws_sitecode,
dbo.USGS_sitefile.state_cd AS usgs_stateCode, dbo.USGS_sitefile.dec_lat_va AS
latitude,
                      dbo.USGS_sitefile.dec_long_va AS longitude
FROM         dbo.usgs_qw_por AS sitePOR INNER JOIN
                      dbo.USGS_sitefile ON sitePOR.SITEID =
dbo.USGS_sitefile.site_id INNER JOIN
```

```
                    dbo.odm_qw_variables AS variables ON
sitePOR.PARAMETER_CD = variables.VariableCode LEFT OUTER JOIN
                    dbo.Units AS Units_1 ON variables.TimeUnitsID =
Units_1.UnitID LEFT OUTER JOIN
                    dbo.Units ON variables.VariableUnitsID =
dbo.Units.UnitID
```

For the daily values view, a variable with the statistical code is created.