AQS Site Registration
Using a Geospatial Interface

Background

At last year's AQS Conference in San Antonio, users identified some issues related to registering new sites in AQS. One of the key problems identified was with the conversion of UTM to latitude/longitude. The AQS conversion software did not always generate the correct coordinates. The new AQS enhancement described in this paper corrects this problem and also provides a new approach for populating several data fields added to AQS when it was re-engineered in 2001, but which have never been populated for existing sites and some new sites.

This proposed AQS enhancement allows users to register a new site and rely upon AQS software to generate input for several data fields. This technical approach was successfully implemented when the tribal modification was implemented a few years ago. The tribal modification allowed tribes to register a site without specifying State and county codes for the site because the technology within AQS was programmed to determine these codes using the user supplied coordinates and geospatial information from the EPA's Envirofacts database to define county boundaries. This proposed enhancement extends this approach to several site level geographic data fields for which geospatial layers are currently available from the EPA's Envirofacts database.

In the future, it could be extended to include other fields as additional geospatial layers become available.

Introduction

This AQS enhancement provides an automatic lookup and/or validation of geopolitical data associated with AQS sites based on the location of the site expressed in terms of either latitude/longitude or UTM coordinates (and associated metadata). The intent is to provide the following capabilities:

1   When a site is created, the user may leave any of the seven (see Supported Geopolitical Fields below) supported geopolitical fields blank (null). The AQS software will then derive the value for those geopolitical fields by using the user supplied locational coordinates and looking up their value in the EPA "master" locational database contained in Envirofacts. If the geopolitical fields cannot be derived from the Envirofacts database (possibly because the location is on a map that has not been loaded into Envirofacts), then a warning will be issued, and the user-provided values will be accepted without validation.

2   If a user supplied value for any of the seven supported geopolitical fields differs from the derived value, the reported value will be rejected and the derived value stored in the AQS database. A warning message will be issued in the Edit Error Detail report. If the user does not agree with the derived values, the coordinates should be verified to ensure that they are not in error. If the coordinates are correct, user support should be notified so that a record of

any differences can be maintained. If there are consistent disagreements with derived values, the process will be re-evaluated.

3   Initially, the geopolitical fields supported by this look up process will be populated for existing sites. Periodically, the values of geopolitical fields will be "refreshed" by re-deriving their values from updated entries in the Envirofacts master locational database. (This will primarily be of use for fields, such as congressional district, that may change over time.)

4   After this enhancement has been implemented, an initial reconciliation will be performed to populate blank fields at existing sites and identify possible differences in populated fields. This initial reconciliation will only be done for sites with a datum of NAD27 because those are the only coordinates that we can assume that both sets (latitude/longitude and UTMs) are equivalent.

The table below lists the number of sites with each datum to give an indication of the number of sites that will be able to be reconciled.

Horizontal Datum	# Sites	# Sites before 01/01/1990	% sites
Unknown	15,043	13,923	73
NAD27	1,913	581	10
NAD83	1,832	853	
WGS84	239	66	
Total	19,027		

We contacted USGS to get some information on maps published by their office. Their office did not begin to change the datum for their published maps from NAD27 until after January 1990. Also, since techniques other than maps were not used before then to determine locations, we feel that the assumption can be made that all sites with an unknown datum that began sampling before January 1990 should be assigned NAD27. Based on the assumption for NAD27, approximately 83% of the sites will be processed through the geospatial look-up process. Submitting agencies would be responsible for updating the other sites based on possible history of which coordinates were used to register sites. The other option would be for agencies to update the coordinates with GPS data.

5   This enhancement also corrects a problem in AQS, which in some cases, converted user-supplied UTM coordinates to incorrect latitude/longitude coordinates (but generally within 200 meters)*.*   Our research has determined that this conversion is dependent on the reported horizontal datum for the site and that the AQS conversion was correct only for site coordinates reported using the NAD27 datum.

6   In addition, the enhancement converts the location in the user's coordinate system (the combination of latitude/longitude or UTM and any valid Horizontal Datum) to a "common" coordinate system (latitude/longitude and Horizontal Datum of WGS84). Both sets of coordinates (the user's coordinates and the derived "common" coordinates) will be stored in the database for the site. This change is intended to facilitate geographical presentation of

retrieved data in a common frame of reference. In the future, AQS will no longer convert latitude/longitude to UTM coordinates and store them in the database. The common frame of reference in latitude/longitude is being created since most GIS work uses this set of coordinates.

7 As mentioned above, the latitude/longitude will be initially converted to the "WGS84" coordinate system for only sites with coordinates reported in NAD27 datum. For all other datum, the submitting agency will be responsible for converting the coordinates to the WGS84 coordinate system based on knowing which coordinates (latitude/longitude or UTM) were originally entered. If the submitting agency does not know which coordinate system was originally used, the only solution may be to determine new coordinates for active sites and inactive sites if the exact location can be determined.

Supported Geopolitical Fields

The following fields can be derived from location using geospatial layers from the Envirofacts database:
1   FIPS State Code
2   FIPS County Code
3   FIPS City Code (five digit)
4   Census Tract and Block (2000 Census)
5   Zip Code
6   Congressional District
7   Urbanized Area
Note: Other fields may be added based on the availability of the relevant geospatial layers and the availability of resources.

Workflow

The proposed new capability will be utilized by the present Load Site and Maintain Site functional capabilities. The workflow for each is described as follows:

Load Site

Insert Operation
1   Site transaction "AA" is submitted with required key fields (state-county-site or tribal code-site). All other currently required fields (including latitude/longitude or UTM coordinates and locational meta data) are still required except for FIPS City Code and Urban Area Code, which are derived from locational data.
2   The user supplied locational data (latitude/longitude or UTM coordinates) are stored as reported with their meta data.
3   The user supplied locational data and horizontal datum are used to convert and store coordinates for the site as latitude/longitude with a horizontal datum of WGS84.
4   Latitude and Longitude (in WGS84) are used to look up all available geopolitical fields.
5   Derived values are used to fill in any missing fields (e.g., Zip Code, Census Tract, etc.)
6   For non-tribal sites, if the State code and/or county code derived from the location are different than the ones supplied by the user, an error is reported in the Edit Error Detail report, and the site is not loaded into the database.

7   If a user supplied value for one of the supported geopolitical fields other than State and county and the value derived from its location differ, then the reported value will be rejected and the derived value stored in the database. A warning message will be issued in the Edit Error Detail report indicating when this happens.
8   The following fields are stored in the AQS database in association with the site:
   a) The derived State Code
   b) The derived County Code
   c) The derived City Code
   d) The derived Census Tract
   e) The derived Census Block
   f) The derived Zip Code
   g) The derived Congressional District
   h) The derived Urbanized Area
   i) The user-provided Latitude
   j) The user-provided Longitude
   k) The user-provided Horizontal Datum
   l) The user provided UTM Northing
   m) The user provided UTM Easting
   n) The user provided UTM Zone
   o) The derived "Standard" Latitude (WGS84)
   p) The derived "Standard" Longitude (WGS84)
   q) The derived Horizontal Datum (WGS84).
Note: All other fields are stored as previously.

Update Operation

1   The site key, either (state-county-site number) or (tribal code-site number) is required, and (as previously) a site must already exist in the database with these key fields.
2   If location (latitude-longitude or UTM) is provided on the update transaction, it is stored for the site. The location and Datum are used to derive the location in latitude/longitude with a horizontal datum of WGS84. All supported fields are derived from the Envirofacts master locational database using the location on the update transaction as described above for Insert.
3   If the location is not provided on the update transaction, the existing location for the site is used to derive any geopolitical fields provided on the update transaction.
4   If the derived state-county differs from that provided on the transaction that includes a location update (for non-tribal sites), the transaction is rejected with an error.
5   If a value for any other field supplied by the user on the update transaction is different than the derived value, a warning is issued, and the derived value is stored in the database.
6   The derived geopolitical fields, the standard latitude-longitude, and the user provided location are stored as described above for inserts.

Maintain Site

Insert

1   A new button is added to the form for "Create Site". The response to the button is a popup window with the fields: Latitude, Longitude, Horizontal Datum and Site Number.
2   If the Horizontal Datum is not WGS84, the location is converted to WGS84.

3   All derived geopolitical fields including the latitude/longitude and horizontal datum of WGS84 are populated and these fields are marked as display only (i.e. the user will not be able to change their values).
4   All other fields (those not derived) are displayed for entry or update as before.

Update

1   When a site is retrieved from the database, the form will display the user supplied location fields (latitude/longitude or UTM and Horizontal Datum), and the "standard" latitude/longitude as display-only, and all geopolitical fields derived from location will be marked as display-only (i.e. the user will not be able to change their values).
2   If the user-supplied location on the form is changed, the form will automatically re-compute all derived geopolitical fields.
3   If this results in a change in State or County, then the change is rejected and an error is displayed.
4   All other fields (those not derived) are displayed for entry or update as before.

Periodic Geo-Data Refresh

1   It is expected that the geospatial layers in the Envirofacts master locational database will be refreshed at irregular intervals, and that the AQS Data Administrator will be notified of these updates.
2   The Data Administrator will execute a procedure for updating geospatially derived data.
3   The Data Administrator will select from a list the field to update.
4   The procedure will execute and update the corresponding field for each site where the standard latitude and longitude are populated.
5   Sites will be processed in Support Agency order. If any fields for a site are changed, an output report is generated with the site-id (state-county-site, or tribe-site), the old value, and the new value. These reports will be used for notification of the support agency.

AQS and Air Quality Data Mart Retrievals

1   From AQS, if a user executes the "Extract Site/Monitor Data" (AMP500) standard report to extract data in input transaction format, the locational information for latitude/longitude, UTM coordinates, and all locational metadata will be written to the transaction.   The user will be responsible for deleting the set of coordinates calculated by AQS before processing the transactions.
2   All site changes (inserts, updates, deletes) written to California dump bucket will continue to be generated out in input transaction format, using the locational information input by the user.
3   All other AQS reports will only display "standard" location coordinates (latitude/longitude, and Horizontal Datum of WGS84).
4   Only the location in standard coordinates will be transferred from AQS to the Data Mart. (Note: This means that GetMonitorData retrieval from the Data Mart may return different data that the corresponding retrieval from AQS.)