Semantic Modeling of Air Quality Database Systems

Author Paige Hodgkinson

ID 922282852

Github: phodgkinson1

San Francisco State University

CS675: Introduction to Database System

Milestone	Date Submitted
1	1/20/2024
2	3/5/2024
3	4/1/2024
4	
5	

Table of Contents

General Project Description

The goal of this project is to model standards for the design of air quality control databases in conceptual and semantic formats in order to implement a relational database for test cases as presented. This database details the many parameters of testing, bias testing, and location information, of various pollutants in the air, and will be compared to the standard of the United States Environmental Protection Agency's AQS or Air Quality System's database. Modeling and improvement of AQS standards are important as embedded testing of air quality will continue to grow and change- especially as new technology is developed which may change which pollutants or factors are measured while the climate destabilizes. We seek to improve about the current standard by improving data fields for less expensive querying, which would benefit both the AQS and the AirNow system which feeds the AQS. A test case is presented for the use of the non-linear Air Quality Index (AQI) established by the EPA, as well as other common uses for this data such climate change research, and soft computing decision cases where living locations are recommended for those with medical conditions.

Low Level Improvements

While relational databases are inherently not scalable to large data sets, we anticipate the use of a distributed SQL database given that older data is not frequently changed or cached. Unlike the EPA table, we will utilizes 10 "qualifier" fields to report the largest combination of QA issues. They are normally give the codes of INFORM, NULL, various Quality Assurance qualifiers, Request Exclusion. We change the dynamic but separating INFORM here into its own parameter because data which is not excluded should still be able to retain a separate, dedicated field for informational purposes, which can be concatenated from QA qualifiers. In addition, as QA data should be considered not reportable until resolved, and not reportable data should be logged in a separate table. As request exclusion indicates the same problem, QA NULL and request exclusion should be logged into a separate data sets. As most contemporary air quality management organizations pull from or add to this aggregate EPA database using the AQCSV file format, their designs tend to contain these same core problems and assets.

Functional Database Requirements

1. Region code (State Code or Tribal ID)
   1.1. A region code is either a state code or a "TT" code
2. Local region code (County Code or Tribal Entity Code)
   2.1. A local region code is either a county code or a tribal entity code
   2.2. A local region code shall be a county code if the region code is a state code
   2.3. A local region code shall be a tribal entity code if the region code is a "TT" code
3. State Code
   3.1. A state code shall be unique.
   3.2. A state code shall be associated with 0 or 1 region Code
   3.3. A state code shall be associated with an entry which does not have a tribal entity code.
   3.4. A state code shall be associated with an entry with a county code as a local region code.
   3.5. A state code shall be associated with many of federal site type.
   3.6. A state code shall be associated with many counties.
   3.7. Each state code shall be associated with only one state name.

4. State Name
   4.1. A state name is a federally recognized name of a state.
   4.2. A state name shall be unique.
   4.3. A state name shall be associated with many of federal site type.
   4.4. Each state name shall be associated with only one state code.
5. Tribal Indicator
   5.1. Shall be input as TT to region code.
   5.2. Shall be associated with many tribal entity codes.
6. County Code
   6.1. A county code shall be unique within the state.
   6.2. A county code shall be associated with an entry which has a state code (conjunct).
   6.3. A county code shall not be associated with an entry which has a tribal entity code.
   6.4. A county shall be associated with only one state.
   6.5. A county code shall be associated with only one county name.
   6.6. A county code shall be associated with 0 or many site numbers.
7. County Name
   7.1. A country name shall be unique within the state.
   7.2. A county code shall be associated with many of federal site type.
   7.3. Each county name shall be associated with only one county code.
8. Tribal Entity Code
   8.1. A tribal entity code shall be unique.
   8.2. A tribal entity code shall be not be associated with a state for region code.
   8.3. A tribal entity code shall be associated with 0 or many site numbers.
9. Tribal Name
   9.1. A tribal name shall be associated with one tribal entity code.
10. Site Number
    10.1. A site number shall be associated with one county code or tribal entity code.
    10.2. A site number shall be associated with 1 or more parameter occurrence codes.
11. Parameter Occurrence Code
    11.1. A parameter occurrence code shall be associated with 1 site number.
    11.2. A parameter occurrence code is unique.
12. Datum type (format type for longitude/latitude)
    12.1. Datum type shall be associated with many longitudes.
    12.2. Datum type shall be associated with many latitudes.
    12.3. Datum type shall be associated with many measurements.
    12.4. Datum types are unique.
13. Geolocation
    13.1. A geolocation shall be associated with many sample measurements.
    13.2. A geolocation shall be associated with only one datum type.
    13.3. A geolocation shall have a longitude of the same datum type.
    13.4. A geolocation shall have a latitude of the same datum type.
    13.5.
14. Parameter Code
    14.1. A parameter code is a unique code for pollutants to be measured.
    14.2. A parameter code shall be associated with many sample measurements.
    14.3. A parameter code shall be associated with only one parameter name.

14.4.

15. Parameter Name
    15.1. A parameter name shall be associated with only one parameter code.
    15.2. A parameter name shall be associated with many sample measurements.
16. Units of Measure
    16.1. Units of measure shall be associated with one method code.
    16.2. Units of measure shall be associated with many sample measurements.
17. Method Type
    17.1. A method type shall be associated with only one method code.
    17.2. A method type shall be associated with many sample measurements.
18. Method Code
    18.1. A method code shall be associated with many units of measure
    18.2. A method code shall be associated with only one method type.
    18.3. A method code shall be associated with many sample measurements.
    18.4. A method code shall be associated with one method detection limit.
    18.5. A method code shall be associated with one uncertainty value.
19. Method Name
    19.1. A method name shall be associated with only one method code.
    19.2. A method name shall be associated with many sample measurements.
20. Method Detection Limit
    20.1. Each method detection limit shall be associated with one method code.
21. Qualifier Inform
    21.1. An inform qualifier is optional for entry.
22. Qualifier 1- QA
    22.1. An QA qualifier is optional for standard entry.
23. Qualifier 2- QA
    23.1. An QA qualifier is optional for standard entry.
24. Qualifier 3- QA
    24.1. An QA qualifier is optional for standard entry.
25. Qualifier 4- QA
    25.1. An QA qualifier is optional for standard entry.
26. Qualifier 5- QA
    26.1. An QA qualifier is optional for standard entry.
27. Qualifier 6- QA
    27.1. An QA qualifier is optional for standard entry.
28. Qualifier 7- QA
    28.1. An QA qualifier is optional for standard entry.
29. Qualifier 8- QA
    29.1. An QA qualifier is optional for standard entry.
30. Qualifier 9- QA
    30.1. An QA qualifier is optional for standard entry.
31. Qualifier- NULL
    31.1. A NULL qualifier is optional.
    31.2. A NULL qualifier is not allowed in standard entry.
32. Qualifier- Request Exclusion
    32.1. An request exclusion qualifier is optional.

32.2. An request exclusion qualifier shall only be associated with entries in the non-reportable data table.

33. Uncertainty
    33.1. An uncertainty value shall be associated with a method code.
34. Sample Measurement
    34.1. A sample measurement shall require only geolocation.
    34.2. A sample measurement shall require a method code.
    34.3. A sample measurement shall be associated with 4 qualifiers for Quality Control.
35. Final Table Entries
    35.1. Each entry sample measurement shall contain three of four qualifier fields for Quality Control.
    35.2. An entry shall require a site type.
    35.3. An entry shall require only one sample measurement.
    35.4. An entry shall contain 0 or one qualifier for Quality Control (Inform)
    35.5. An entry shall contain a date of last change.
    35.6. An entry shall derive a field for GMT date from local date.
    35.7. An entry shall derive a field of GMT time from local time.
36. Non-reportable entry
    36.1. The NR entries shall contain a NULL or a request exlusion type qualifier.
37. Database structure
    37.1. The DB shall support concurrent data entry.

Non-Functional Database Requirements

1. Querying the database from outside the organization shall be permitted only through the use of an API developed for semi-public access.
2. Submitting to AQS (Agency API)
   2.1. Accounts
      2.1.1. A submitting agency shall have many accounts.
      2.1.2. Each account shall be associated with only one agency.
      2.1.3. Each account shall be associated with only one email.
      2.1.4. Each email shall be associated with only one account.
      2.1.5. Each account shall have at most one current login.
   2.2. Sites
      2.2.1. Each agency shall have one site list of 0 or many sites.
      2.2.2. Sites shall be associated with one and only one agency.
      2.2.3. Each account shall only have access to the sites of its agency.
      2.2.4. Each account may add or remove sites of its agency.
      2.2.5. An agency shall have 0 or more sites.
      2.2.6. Sites shall have 1 or more measuring devices.
      2.2.7. Sites with more than one measuring device per pollutant type shall have one POC (parameter occurrence code) per device
      2.2.8. Agencies shall submit 0 or more AQCSV data files to the dedicated FTP server.
   1.4 AQCSV files
      1.1.1. All parameters shall be submitted in the correct order of parameters
      1.1.2. All file must be submitted as comma separated values CSV
      1.1.3. Each parameter code must have its own line in CSV
      1.1.4. File Requirements (not per parameter code) may be associated with only one line of CSV
         1.1.4.1. Only the first line of CSV per file may contain file requirements
         1.1.4.2. Each file shall have only one longitude
         1.1.4.3. Each file shall have only one latitude
         1.1.4.4. Each file shall have only one site ID
         1.1.4.5. Each file shall contain 0 or one county code
         1.1.4.6. Each file shall contain 0 or one state code.
         1.1.4.7. Each file shall contain 0 or 1 tribal ID.
         1.1.4.8. Each file shall contain 0 or 1 tribal entity code.
         1.1.4.9. Each file shall contain 0 or 1 tribal entity name.
         1.1.4.10. Each file shall contain 0 or 1 state name.
         1.1.4.11. Each file shall that contains a state code shall contain a county code.
         1.1.4.12. Each file shall that contains a tribal ID shall contain a tribal code.
         1.1.4.13. Each file that contains a state code shall contain a state name.
         1.1.4.14. Each file shall contain 0 or 1 county code.
         1.1.4.15. Each file that contains a county code shall contain a county name.
         1.1.4.16. Each file shall contain either a state code or a tribal ID.
         1.1.4.17. Each file shall contain 0 or more site number.
         1.1.4.18. Each file shall contain 0 or 1 POC or parameter occurrence code.

1.1.4.19. Each file shall contain only one datum type (for longitude and latitude).
1.1.4.20. Each file shall contain one or many parameter code
1.1.4.21. Each file shall contain a parameter name associated with each parameter code
1.1.4.22. Each file shall contain only one local date
1.1.4.23. Each file shall contain only one local time
1.1.5. Per line/per solo parameter code or POC per parameter code
1.1.5.1. Each parameter code shall be associated with 1 or more POC
1.1.5.2. Parameter codes with POC values greater than 1 maybe be associated with more than one measurement per code.
1.1.5.3. Each parameter code with only one POC must be associated with only one measurement.
1.1.5.4. Each POC per parameter code shall contain one or more sample measurement.
1.1.5.5. Each sample measurement shall be associated with one parameter code POC.
1.1.5.6. One measurement shall be associated with each parameter code POC.
1.1.6. Parameter Codes (each line of CSV)
1.1.6.1. Each parameter code shall be associated with only one unit of measure.
1.1.6.2. Units of measure may be associated with multiple parameter codes.
1.1.6.3. Each POC per parameter code per file shall be associated with one uncertainty value.
1.1.6.4. Each POC per parameter code per file shall contain between 0 and 4 qualifiers.
1.1.6.5. Each POC per parameter code per file shall be associated with one method type.
1.1.6.6. Method types shall be associated with 0 or more parameter codes.
1.1.6.7. Each POC per parameter code per file shall contain only one method code.
1.1.6.8. Each method code per file shall be associated with 0 or many POC per parameter codes.
1.1.6.9. Each method code per file shall be associated with only one method types.
1.1.6.10. Each method type file shall be associated with only one method code.

1.2. FTP Server
1.2.1.
1.2.2. File backlog shall be sent via FTP server to an intermediary QC check table.
1.2.3. Backlog sending shall occur at a set time or time interval.
1.2.4. FTP server shall allow a maximum number of files per transfer.

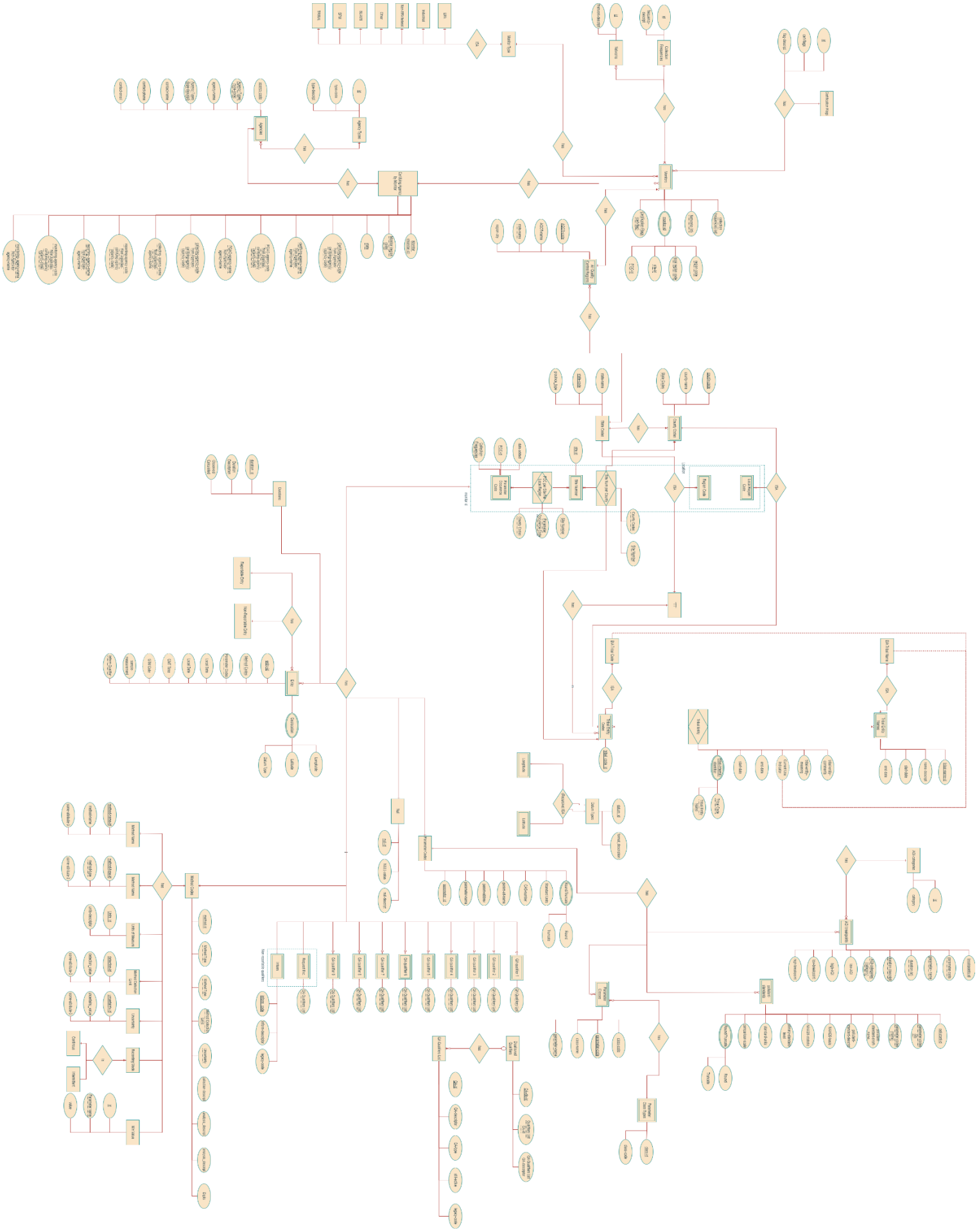
1.3. Intermediary Table
1.3.1. The database shall have a Quality Control intermediary table.
1.3.2. The intermediary shall accept many files from the server.
1.3.3. The intermediary shall check the files for proper formatting and data types.
1.3.4. The intermediary will quality control each file to accept or reject it.
1.3.5. Each file will be accepted or rejected.
1.3.6. Accepted files will be parsed into individual entries
1.3.7. Parsing Entries

1.3.7.1. Each line from the CSV which contains a qualifier not equal to "INFORM" shall be send as one entry to the database primary table
1.3.7.2. Each line from the CSV which contains a qualifier not equal to "INFORM" or no qualifier shall be send as one entry to the database primary table
1.3.7.3. Each entry shall inherit the required file parameters from the first CSV line

2. User facing API

Entity Relationship Diagram

Redone- April 1st 2023 V2



Entity Description

1) State Codes
    a) state-code: key, numeric
b) state-name: attribute, alphanumeric
    c) some-attribute-3: attribute, alphanumeric
2) County Codes
    a) county-code: key, numeric
    b) state-code: weak key, numeric
    c) county-name: attribute, alphanumeric
    d) state-name: derived attribute from State Codes, alphanumeric
3) Tribal Entity Code
    a) tribal-entity-code: key, alphanumeric
    b) start-date: attribute, date
    c) end-date: attribute, date
4) Tribal Entity Names
    a) tribal-entity-name: key, alphanumeric
    b) start-date: attribute, date
    c) end-date: attribute, date
5) Site Number
    a) site-id: key, numeric
    b) county or tribal entity code: compound attribute, alphanumeric
    c) some-attribute-3: attribute, alphanumeric
6) Parameter Occurrence Codes
    a) POC-id: key, numeric
    b) site-id: weak key, numeric
    c) some-attribute-3: attribute, alphanumeric
7) TT
    a) TT-id: key, numeric
    b) TT-descript: attribute, alphanumeric
    c) some-attribute-3: attribute, alphanumeric
8) Tribal Entity Group
    a) tribal-entity-group-id: key, alphanumeric
    b) tribal-group-name: attribute compound (from tribal entity name and tribal entity code), numeric
    c) Internal Tribal ID: attribute, alphanumeric
    d) Current use indicator: derived attribute, alphanumeric
    e) Start date: attribute, date
    f) End date: attribute, date
    g) Tribal entity description: attribute, alphanumeric
    h) Tribal entity comments: attribute, alphanumeric
9) Datum Types
    a) datum-id: key, numeric
    b) format-description: attribute, alphanumeric
    c) some-attribute-3: attribute, alphanumeric
10) Parameter Codes

a) parameter-id: key, numeric
b) parameter-name: attribute, alphanumeric
c) some-attribute-3: attribute, alphanumeric

11) Method Codes
a) method-id: key, numeric
b) Method Name: weak key, numeric
c) Unites of Measurement: weak key, numeric
d) Detection Limit: weak key, numeric
e) Uncertainty: weak key, numeric

12) Method Name
a) method-name-id: key, numeric
b) method-name: attribute, alphanumeric
c) some-attribute-3: attribute, alphanumeric

13) Method Type
a) method-type-id: key, numeric
b) method-type: attribute, alphanumeric
c) some-attribute-3: attribute, alphanumeric

14) Units of Measurement
a) units-id: key, numeric
b) units-descriptor: attribute, alphanumeric
c) some-attribute-3: attribute, alphanumeric

15) Detection Limit
a) detection-id: key, numeric
b) detection-value: attribute, numeric
c) some-attribute-3: attribute, alphanumeric

16) Uncertainty
a) uncertainty-id: key, numeric
b) uncertainty-value: attribute, numeric
c) some-attribute-3: attribute, alphanumeric

17) Non-reportable entry
a) nre-id: key, numeric
b) null-id: weak key, numeric
c) reqex-id: weak key, numeric

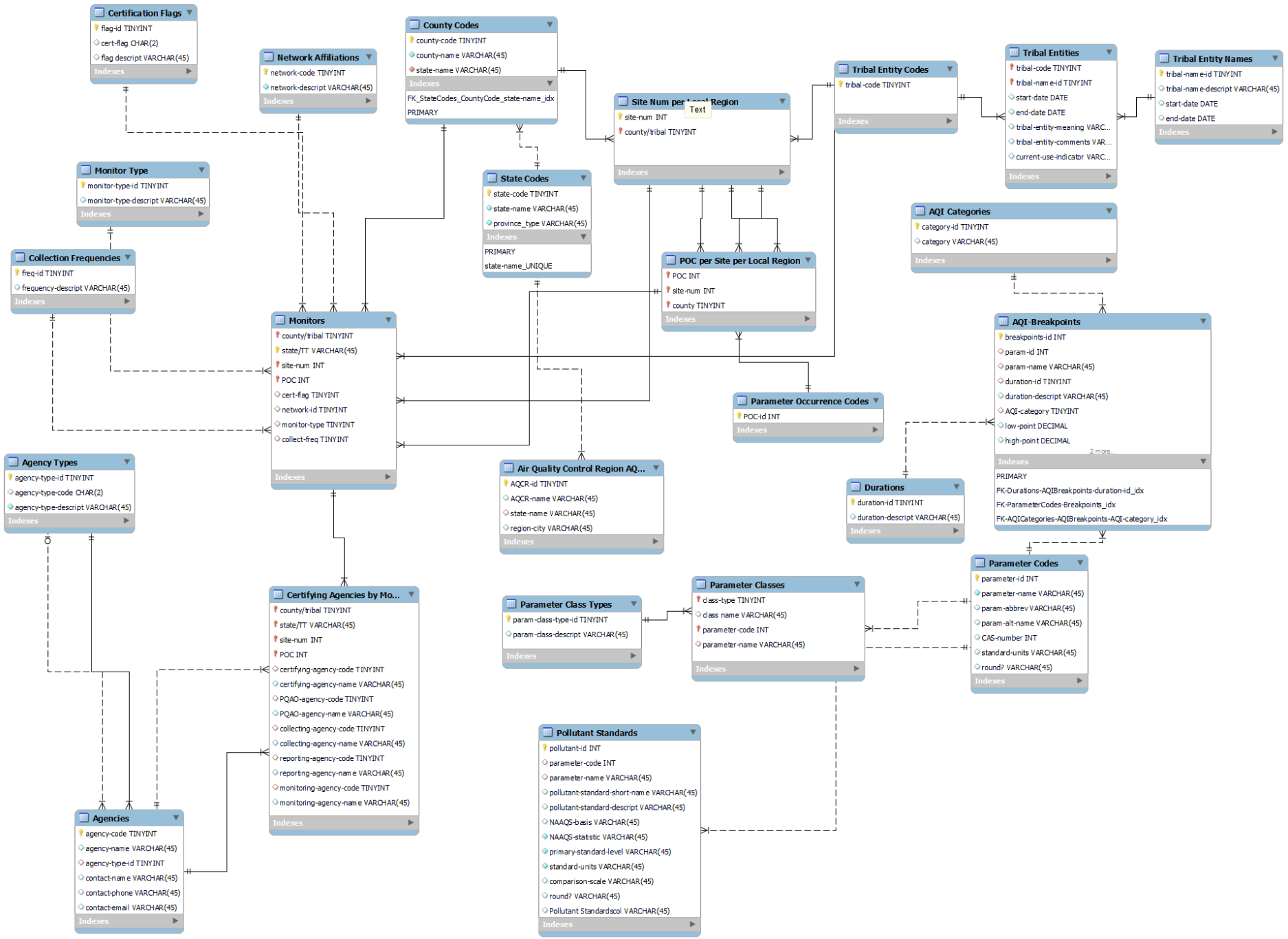
18) Reportable entry
a) QA Qualifiers List (inform): weak key, alphanumeric, null okay
b) QA Qualifiers List: weak key, alphanumeric, null okay
c) QA Qualifiers List: weak key, alphanumeric, null okay
d) QA Qualifiers List: weak key, alphanumeric, null okay
e) QA Qualifiers List: weak key, alphanumeric, null okay
f) QA Qualifiers List: weak key, alphanumeric, null okay
g) QA Qualifiers List: weak key, alphanumeric, null okay
h) QA Qualifiers List: weak key, alphanumeric, null okay
i) QA Qualifiers List: weak key, alphanumeric, null okay
j) QA Qualifiers List: weak key, alphanumeric, null okay

19) QA Qualifiers List
a) QA-id: key, alphanumeric

b) QA-descriptor: attribute, numeric
c) QA-type: weak key, alphanumeric
d) still-active: attribute, Boolean
e) legacy-code: attribute, alphanumeric

20) Inform Qualifier
    a) Inform-code: key, numeric
    b) Inform-descriptor: attribute, alphanumeric
    c) some-attribute-3: attribute, alphanumeric
21) Entry
    a) Location (aggregate)
        i) region-code: attribute, numeric
        ii) local-region-code: attribute, numeric
        iii) site-id: weak key, numeric
        iv) POC-id: attribute derived from Parameter Occurrence Codes, numeric
    b) Geolocation (compound)
        i) datum_id: weak key, numeric
        ii) longitude: weak key, alphanumeric
        iii) latitude: weak key, alphanumeric
    c) entry-id: key, numeric
    d) parameter-id: weak key, numeric
    e) method-id: weak key, numeric
    f) local-time: attribute, time
    g) local-date: attribute, date
    h) GMT-date: derived attribute, date
    i) GMT-time: derived attribute, time
    j) sample-measurement: attribute, numeric

Section VIII: Entity Establishment Relationship Diagram (EER)


Certification Flags
flag-id TINYINT
cert-flag CHAR(2)
flag descript VARCHAR(45)
Indexes
Network Affiliations
network-code TINYINT
network-descript VARCHAR(45)
Indexes
County Codes
county-code TINYINT
county-name VARCHAR(45)
state-name VARCHAR(45)
Indexes
FK_StateCodes_CountyCode_state-name_idx
PRIMARY
Site Num per Local Region
Text
site-num INT
county/tribal TINYINT
Indexes
Tribal Entity Codes
tribal-code TINYINT
Indexes
Tribal Entities
tribal-code TINYINT
tribal-name-id TINYINT
start-date DATE
end-date DATE
tribal-entity-meaning VARC...
tribal-entity-comments VAR...
current-use-indicator VARC...
Indexes
Tribal Entity Names
tribal-name-id TINYINT
tribal-name-descript VARCHAR(45)
start-date DATE
end-date DATE
Indexes
Monitor Type
monitor-type-id TINYINT
monitor-type-descript VARCHAR(45)
Indexes
State Codes
state-code TINYINT
state-name VARCHAR(45)
province_type VARCHAR(45)
Indexes
PRIMARY
state-name_UNIQUE
AQI Categories
category-id TINYINT
category VARCHAR(45)
Indexes
Collection Frequencies
freq-id TINYINT
frequency-descript VARCHAR(45)
Indexes
POC per Site per Local Region
POC INT
site-num INT
county TINYINT
Indexes
Monitors
county/tribal TINYINT
state/TT VARCHAR(45)
site-num INT
POC INT
cert-flag TINYINT
network-id TINYINT
monitor-type TINYINT
collect-freq TINYINT
Indexes
Parameter Occurrence Codes
POC-id INT
Indexes
AQI-Breakpoints
breakpoints-id INT
param-id INT
param-name VARCHAR(45)
duration-id TINYINT
duration-descript VARCHAR(45)
AQI-category TINYINT
low-point DECIMAL
high-point DECIMAL
2 more...
Indexes
PRIMARY
FK-Durations-AQIBreakpoints-duration-id_idx
FK-ParameterCodes-Breakpoints_idx
FK-AQICategories-AQIBreakpoints-AQI-category_idx
Agency Types
agency-type-id TINYINT
agency-type-code CHAR(2)
agency-type-descript VARCHAR(45)
Indexes
Air Quality Control Region AQ...
AQCR-id TINYINT
AQCR-name VARCHAR(45)
state-name VARCHAR(45)
region-city VARCHAR(45)
Indexes
Durations
duration-id TINYINT
duration-descript VARCHAR(45)
Indexes
Parameter Codes
parameter-id INT
parameter-name VARCHAR(45)
param-abbrev VARCHAR(45)
param-alt-name VARCHAR(45)
CAS-number INT
standard-units VARCHAR(45)
round? VARCHAR(45)
Indexes
Certifying Agencies by Mo...
county/tribal TINYINT
state/TT VARCHAR(45)
site-num INT
POC INT
certifying-agency-code TINYINT
certifying-agency-name VARCHAR(45)
PQAO-agency-code TINYINT
PQAO-agency-name VARCHAR(45)
collecting-agency-code TINYINT
collecting-agency-name VARCHAR(45)
reporting-agency-code TINYINT
reporting-agency-name VARCHAR(45)
monitoring-agency-code TINYINT
monitoring-agency-name VARCHAR(45)
Indexes
Parameter Class Types
param-class-type-id TINYINT
param-class-descript VARCHAR(45)
Indexes
Parameter Classes
class-type TINYINT
class name VARCHAR(45)
parameter-code INT
parameter-name VARCHAR(45)
Indexes
Pollutant Standards
pollutant-id INT
parameter-code INT
parameter-name VARCHAR(45)
pollutant-standard-short-name VARCHAR(45)
pollutant-standard-descript VARCHAR(45)
NAAQS-basis VARCHAR(45)
NAAQS-statistic VARCHAR(45)
primary-standard-level VARCHAR(45)
standard-units VARCHAR(45)
comparison-scale VARCHAR(45)
round? VARCHAR(45)
Pollutant Standardscol VARCHAR(45)
Indexes
Agencies
agency-code TINYINT
agency-name VARCHAR(45)
agency-type-id TINYINT
contact-name VARCHAR(45)
contact-phone VARCHAR(45)
contact-email VARCHAR(45)
Indexes