AQS Database Structure

Presented by
Jonathan Miller, USEPA
RTP, NC

Section Objectives

- Understand the Basic Terminology of a Database
- Learn What Types of Data are Contained within the AQS Database
- Understand What Makes a Given Record Unique Within the Database
- Learn the Required Components of Each Type of Data

Database Terminology

- Table
  - A Structured Storage Unit within a Database
  - Contains one or more Records (Rows)
  - Each Record has the same type of information, defined by Columns
  - Most Tables Use One or More Columns to Uniquely Identify Records. These are Called *UNIQUE KEYS* or *PRIMARY KEYS*

Table Definition Example

The ATTENDEES table contains information about people attending the AQS training class. Identify the following elements in this example:

1) Records
2) Columns
3) What Might Make up the Unique Key?

Attendees

NAME	Phone #	E-Mail	GenderCode
Joe Cool	123-456-7890	joecool@yahoo.com	M
Sally Forth	890-123-4567	sallyforth@msn.net	F
Mike RaFone	222-333-4444	mikeraphone@mailme.com	M

Database Terminology

- Foreign Key
  - A column on a table that references the PRIMARY KEY of another table.
  - These keys are used to “walk” from one table to another
  - Defines the “Parent” / “Child” Relationship Between Tables
    - Parent Tables have the Primary Key (PK)
    - Child Records Use Foreign Keys to Reference the Parent Record
  - Ensures Referential Integrity Between Tables

Primary / Foreign Key Example

Genders

GENDER CODE	GENDER DESCRIPTION
M	MALE
F	FEMALE

PK of GENDERS

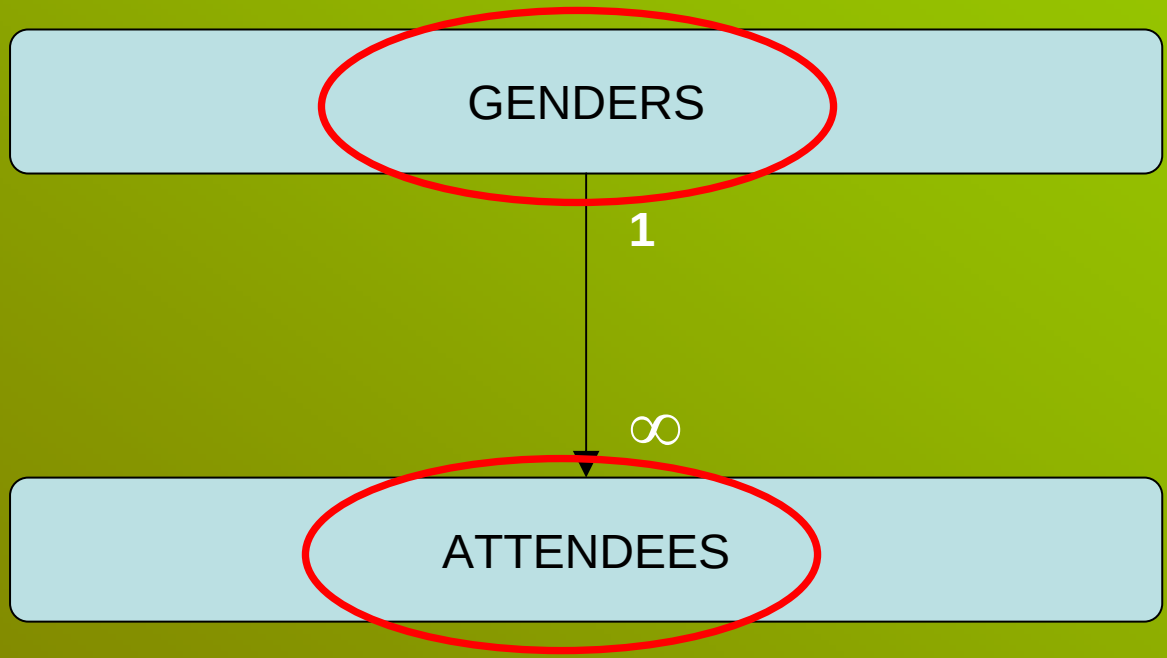
Attendees

NAME	Phone #	E-Mail	GenderCode
Joe Cool	123-456-7890	joecool@yahoo.com	M
Sally Forth	890-123-4567	sallyforth@msn.net	F
Mike RaFone	222-333-4444	mikeraphone@mailme.com	M

FK to GENDERS

What Would the Data Model Look Like?

EACH GENDER CAN APPLY TO MULTIPLE ATTENDEES.



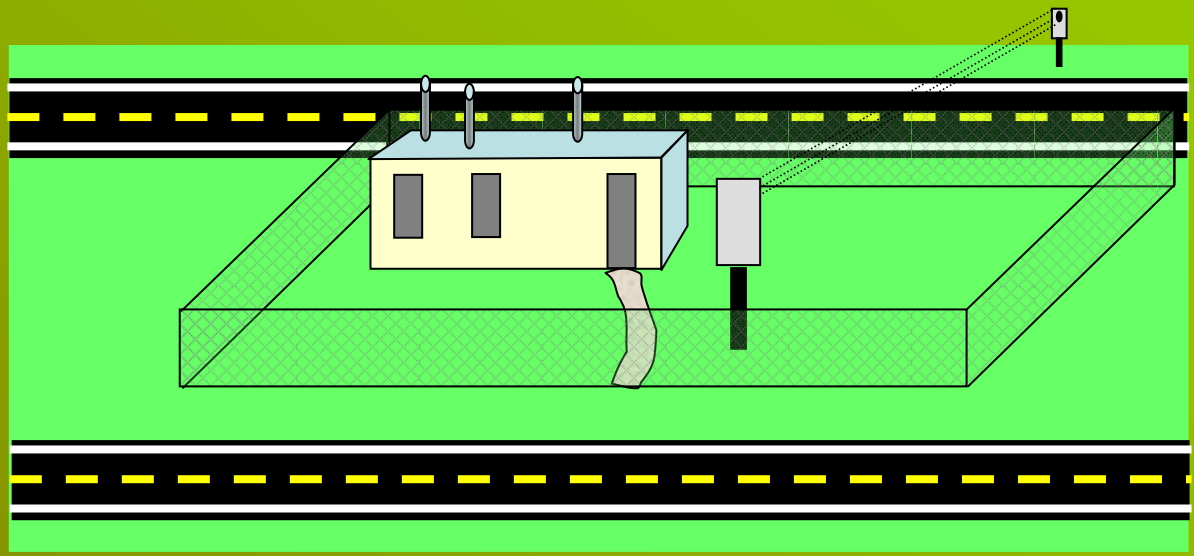
EACH ATTENDEE IS ASSIGNED ONE GENDER.

SITES

Identifies WHERE the site is located:

- Latitude / Longitude
- Tribal Area
- State
- County
- City
- Air Quality Control Region

SITES



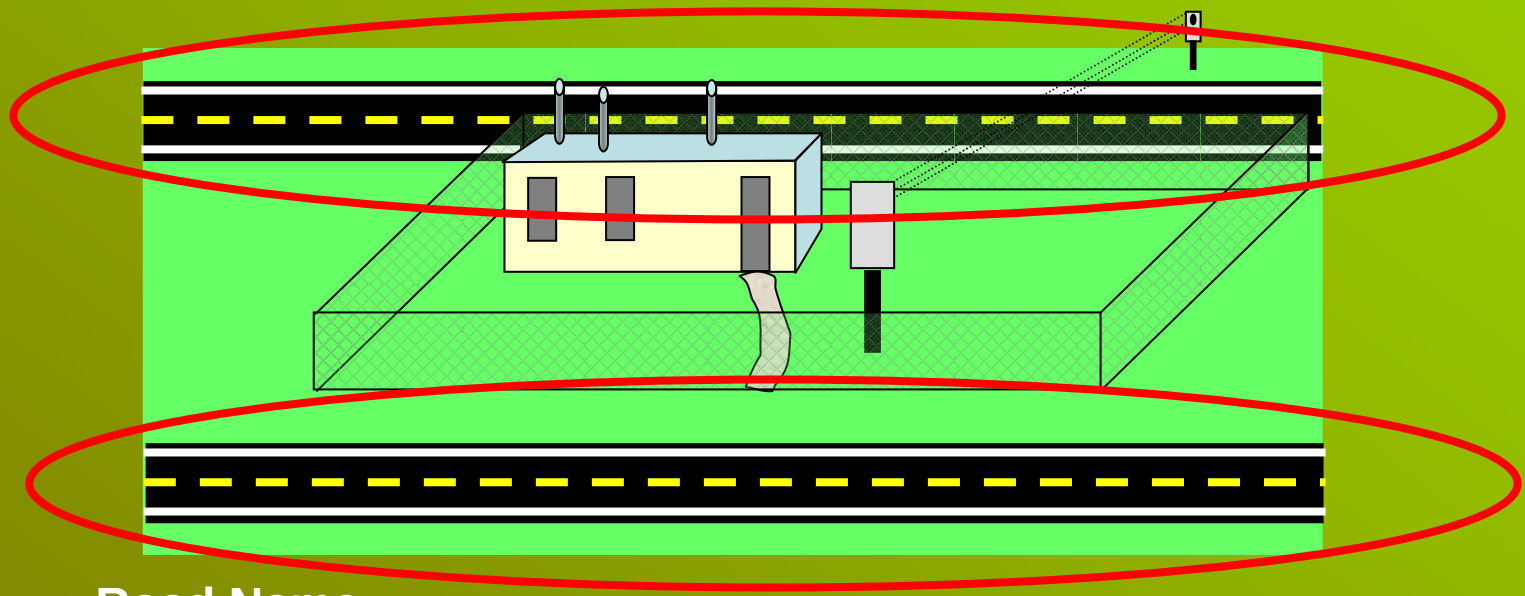
Provides Identifying Information

- Site ID
- Local Site ID
- Local Name

SITES

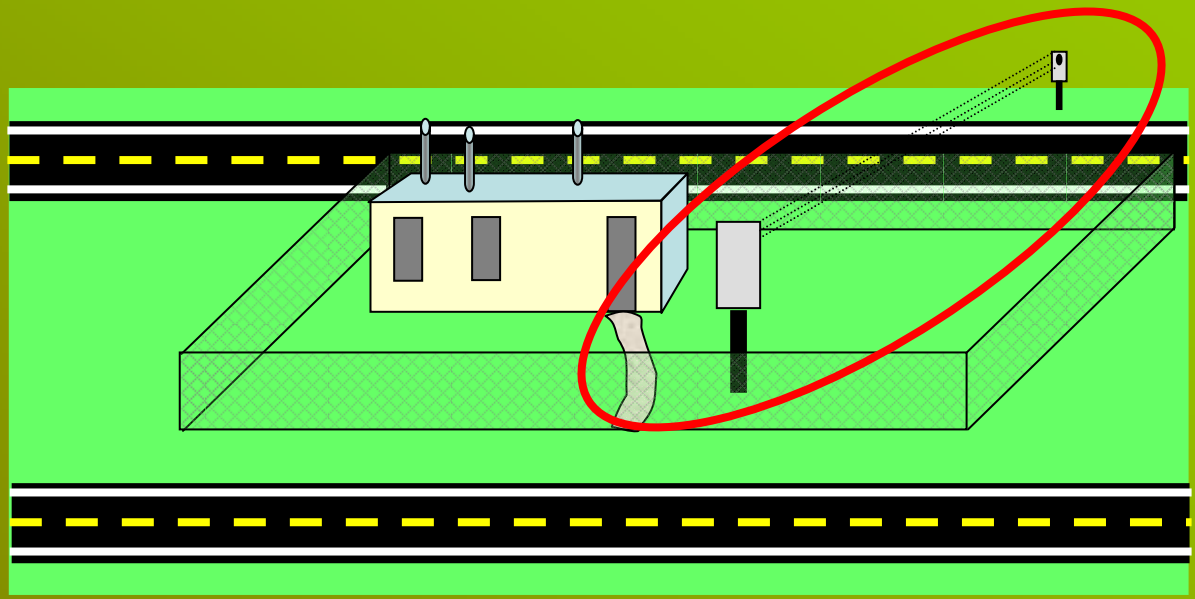
- UNIQUE KEY
  - State Code, County Code, & Site ID
  - Alternate Key: Tribal Area & Site ID
  - This Combination is Called the AQS SITE ID
- Child Records
  - TANGENT ROADS
  - OPEN PATHS
  - MONITORS
  - AGENCY ROLES (SUPPORTING)

Tangent Roads & Open Paths



Road Name
Type of Road (Local, Highway, Interstate, etc)
Daily Traffic Count
Year Traffic Count Obtained
Source of the Traffic Count
Direction from Site to Street

Tangent Roads & Open Paths

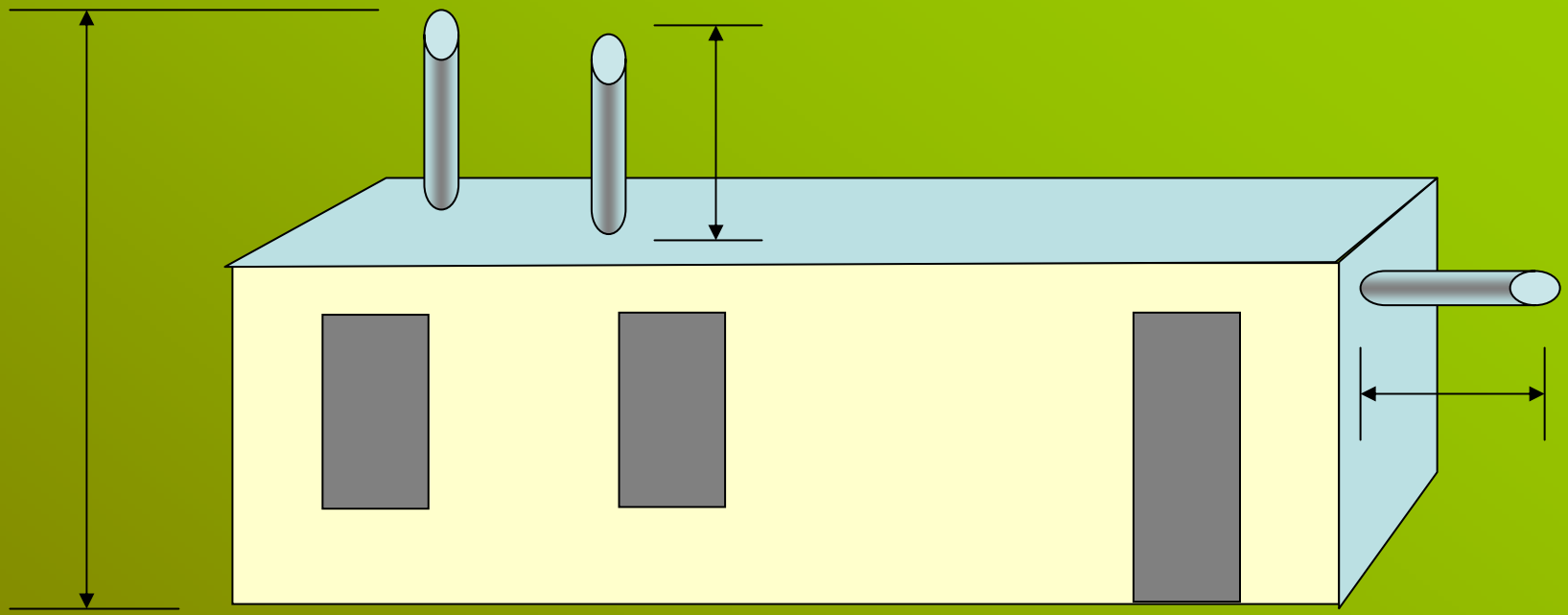


Transmitter Height
Receiver Height
Land Use Under Path
Path Length
Direction From Receiver to Transmitter
Minimum Height
Maximum Height

Tangent Roads & Open Paths

- OPEN PATHS
  - UNIQUE KEY
    - AQS SITE ID + Open Path Number
  - Child Records: NONE
- TANGENT ROADS
  - UNIQUE KEY
    - AQS SITE ID + Tangent Street Number
  - Child Records: NONE

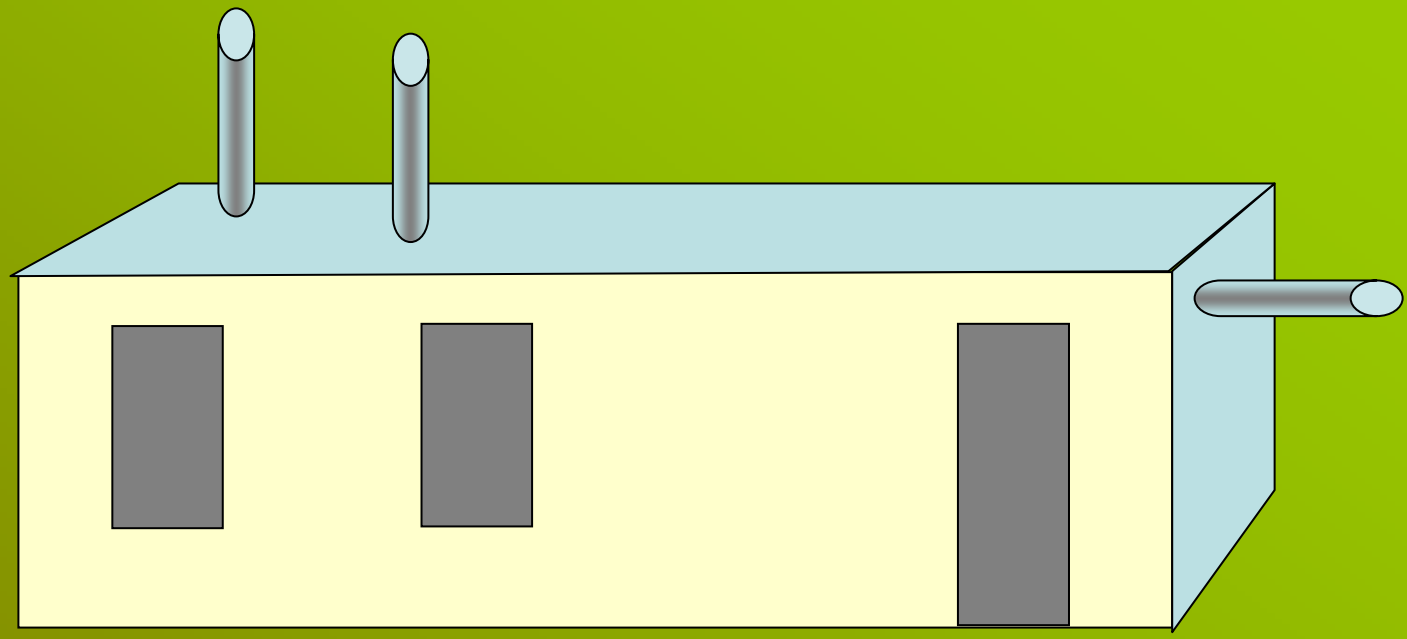
MONITORS



Probe Information

- **What is Being Measured?**
- **Probe Height**
- **Probe Vertical Distance**
- **Probe Horizontal Distance**
- **Is the Probe Obstructed?**

MONITORS



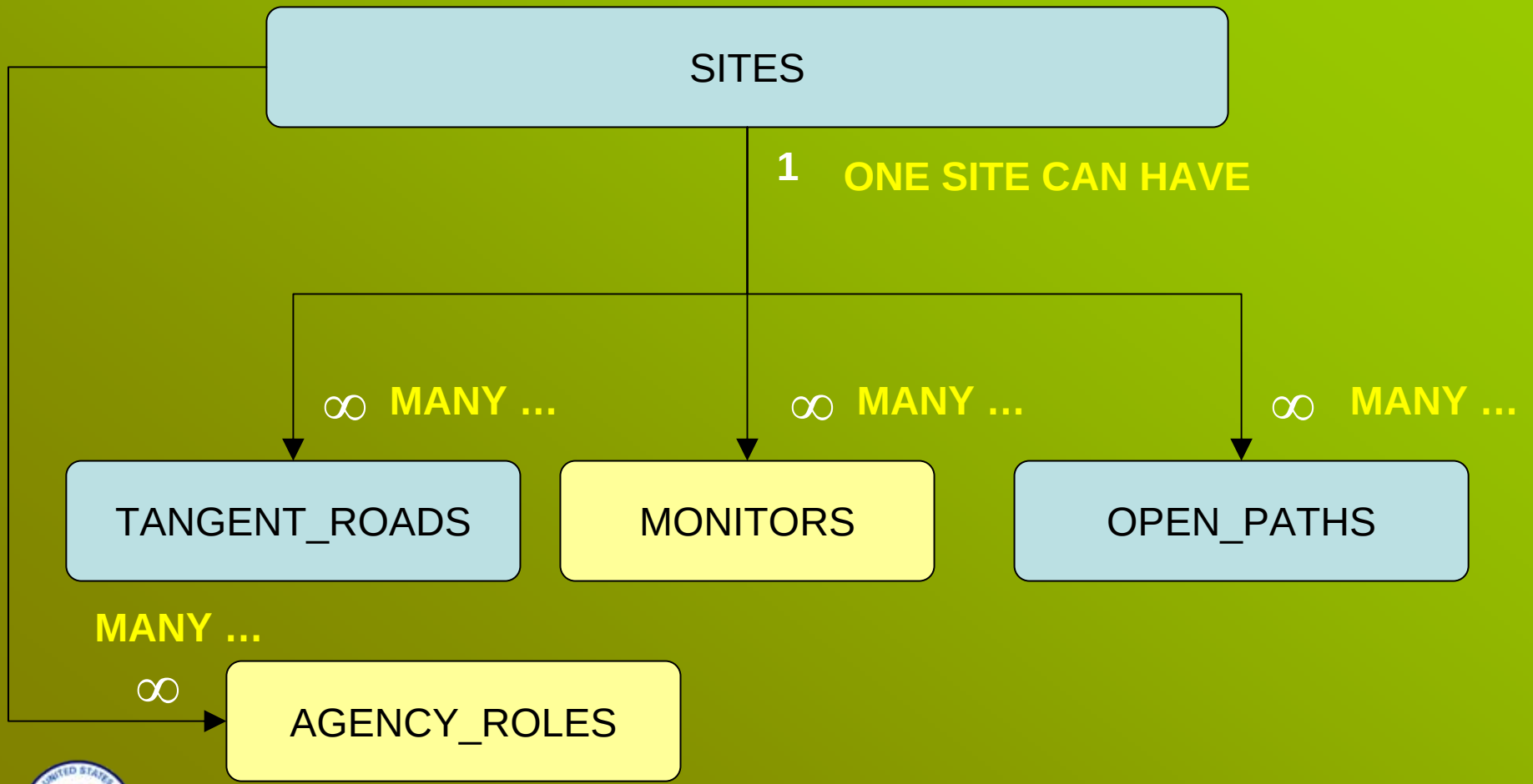
What Impacts the Data from This Monitor?

- **Dominant Source**
- **Measurement Scale**
- **Project Classification**
- **Sample Residence Time**

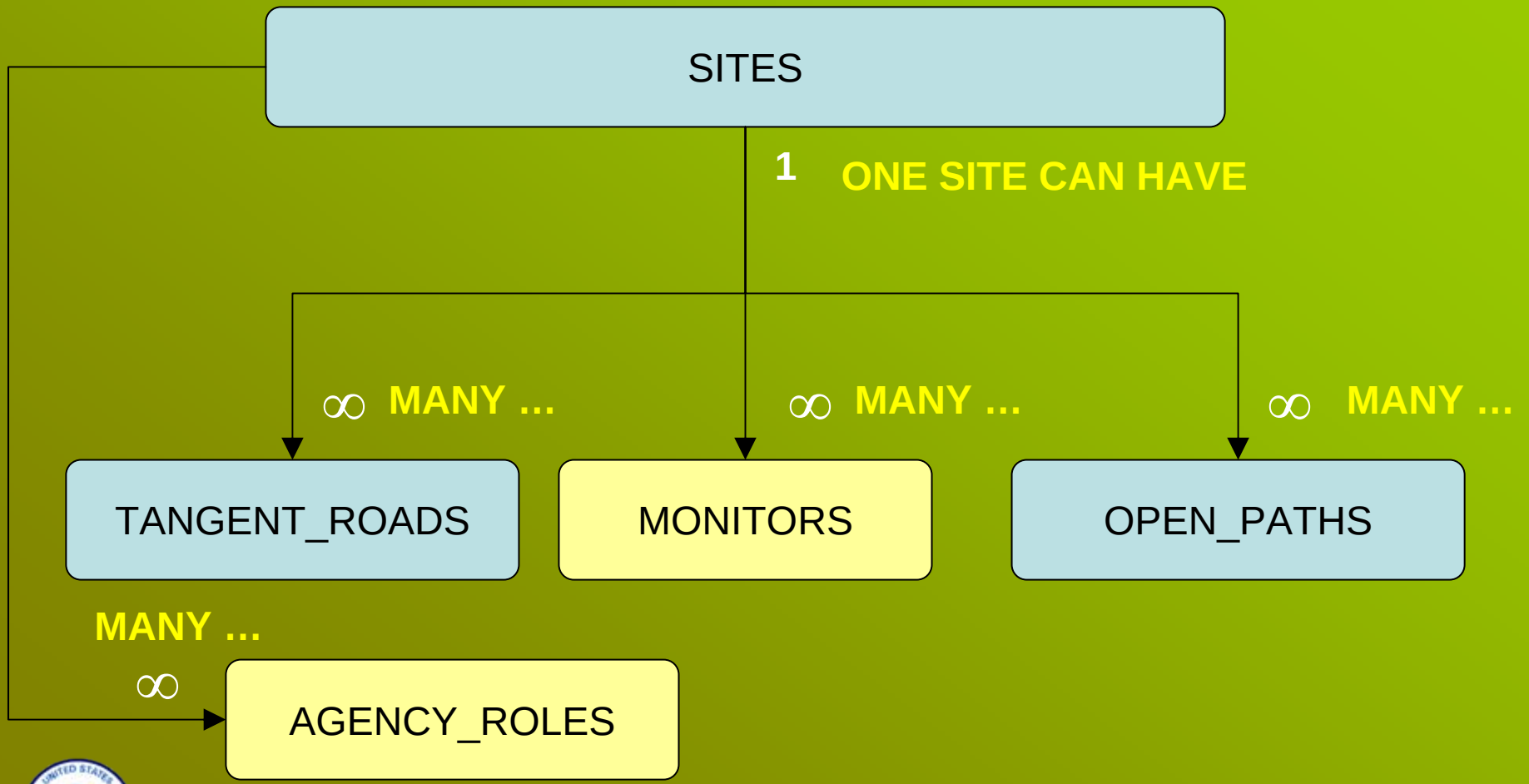
MONITORS

- UNIQUE KEY
  - AQS SITE ID + Parameter Code + Parameter Occurrence Code (POC)
    - This Combination is Called the AQS MONITOR ID
    - An AQS Monitor Tracks what is being collected on a pollutant by pollutant basis. Does Not Reflect Specific Instrumentation
    - POC is a identification number distinguishing multiple instruments that may measure the same pollutant

What's the Model Look Like So Far?



Take a Closer Look at MONITORS

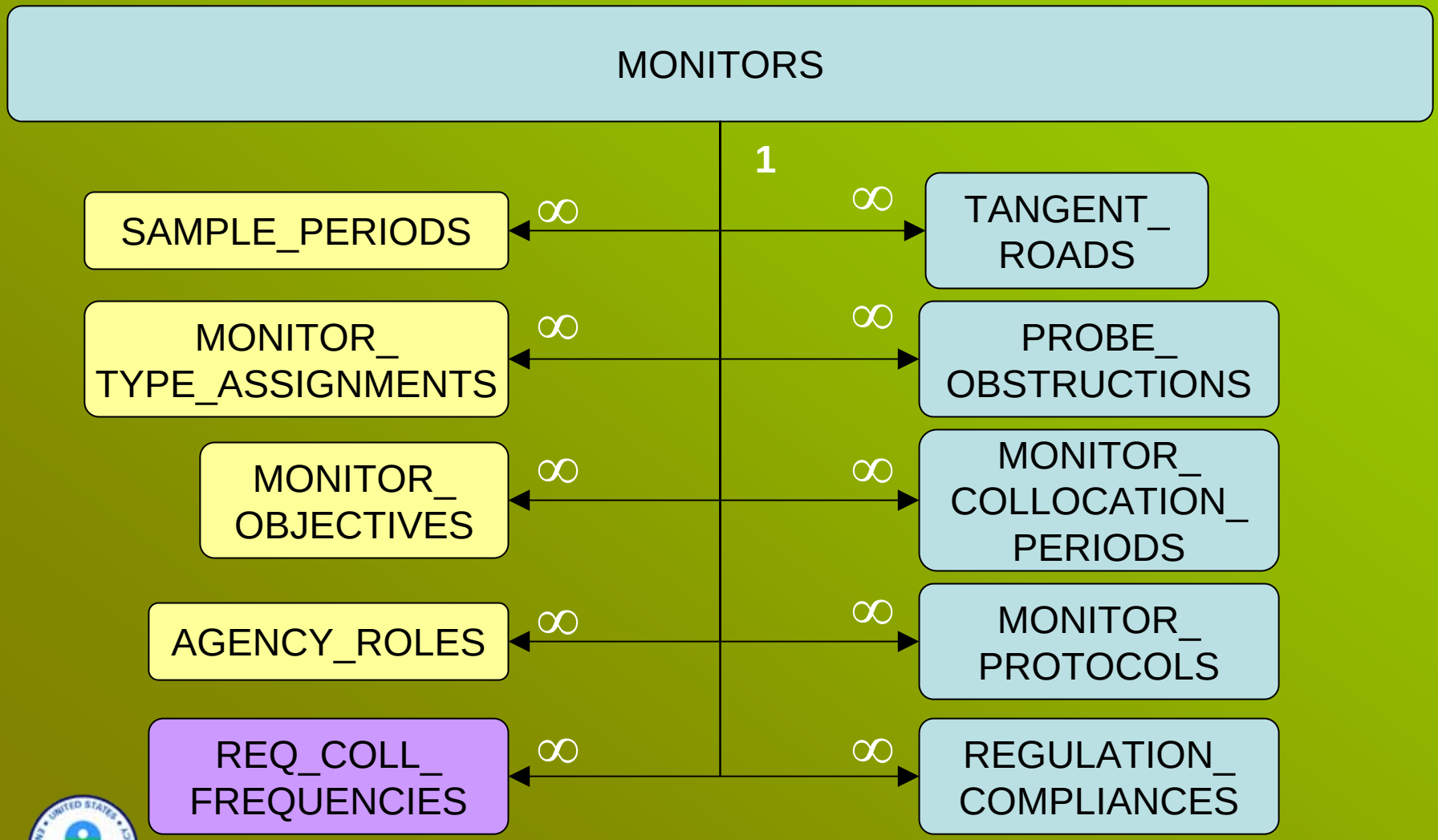


MONITORS Information

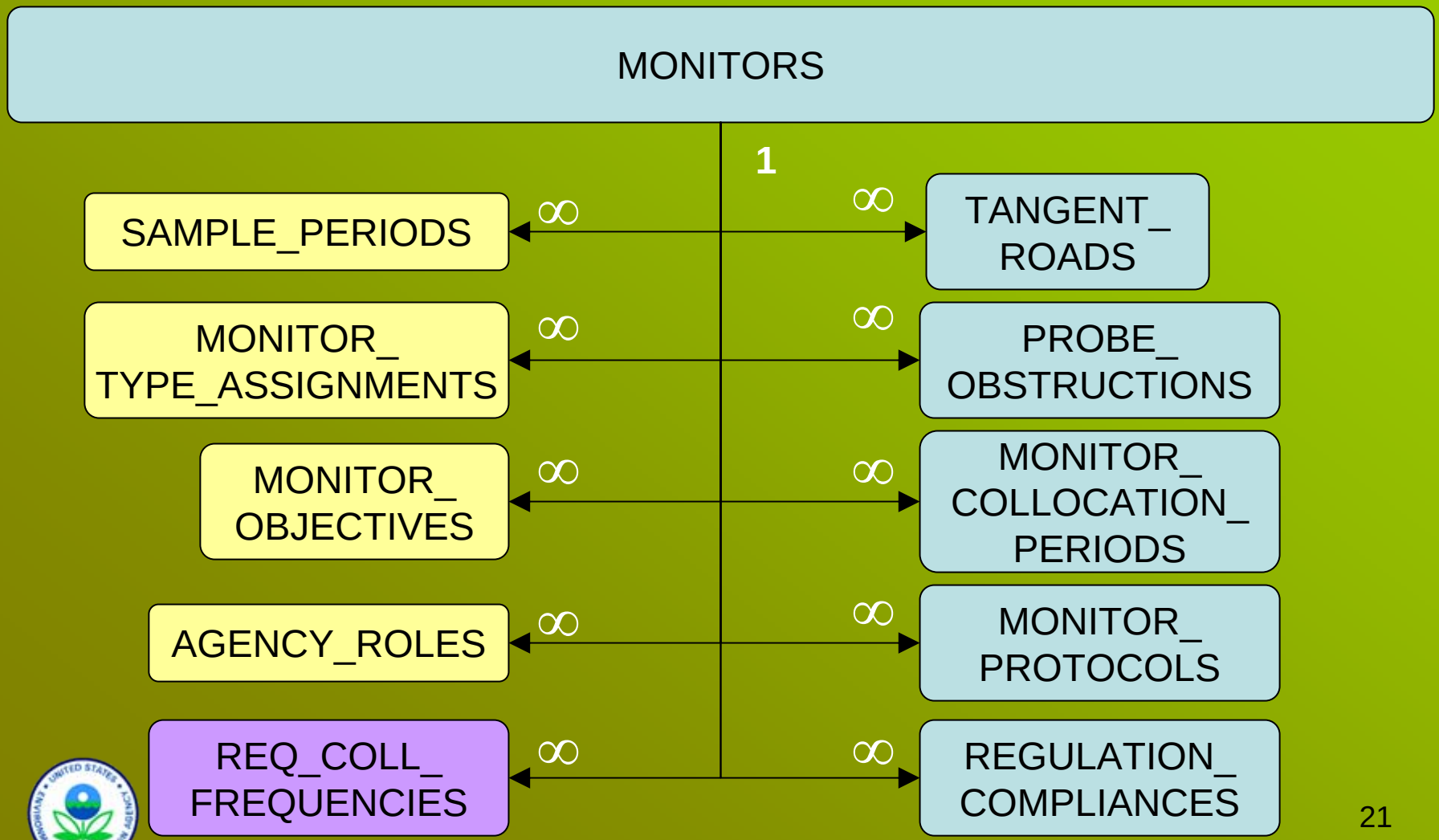
- Categories of Monitor Information
  - Regulatory
    - Agency Roles (REPORTING, COLLECTING, ANALYZING)[1]
    - Monitor Types (Networks) [1]
    - Monitoring Objectives[1]
    - Regulation Compliances
  - What May Impact the Monitor's Data
    - Probe Obstructions
    - Monitor Tangent Roads
  - How the Monitor Collects Information
    - Sample Periods [1]
    - Monitor Protocols
    - Required Collection Frequencies[2]
    - Monitor Collocation Periods

1→ Required Data Types **2→ Conditionally Required Data Types**

MONITORS Data Model



A Closer Look at MONITOR_PROTOCOLS



What is MONITOR_PROTOCOLS

- Contains Specific Information as to How the data was Collected
  - Sampling Methodology
  - Sample Duration
  - Units of Measure
  - Collection Frequency
  - Alternate Minimum Detection Limit
  - Composite Type (If Doing COMPOSITE type of Sampling)

MONITOR_PROTOCOLS

- Required Information For all Types of Data Input Into the System
  - Raw Data
    - “Simple” Raw Data
    - Composite Raw Data
    - Blanks Data
  - Precision Data
  - Accuracy Data

MONITOR_PROTOCOLS Cont.

- Unique Key
  - Monitor ID + Unique Combination of:
    - Method
    - Duration
    - Unit of Measure
    - Collection Frequency
    - Composite Type
    - Alternate MDL
- Primary Key
  - AQS Monitor ID + Monitor Protocol Sequence Number

Raw Data

- Individual Monitoring Points
- Most of the Data in AQS is of this type (1,000,000,000+ values)
- Can Have Associated “Qualifier” Information
  - Quality Assurance Issues Occurred with the Sample
  - Natural Event Occurring which affected the Sample
  - Exceptional Event Occurring which affected the Sample
- Unique Key
  - Monitor ID
  - Date / Time of Collection

Precision & Accuracy Data

- Auditing Information
- Unique Key
  - Monitor ID
  - Date of Collection
  - Precision or Accuracy ID (Sequence Number)

Summary Data

- System Generates Multiple “Time Levels” of Raw Data Summaries
  - Annual Summaries
  - Quarterly Summaries
  - Daily Summaries
  - NAAQS Averages
- These Summaries Do Not Carry Monitor_Protocol Information (Methods)
- Unique Key
  - Monitor
  - Time Frame (Year, Quarter, Day)
  - Duration
  - Exceptional Data Type Code (0 – 7)

Exceptional Data Types (EDT)

Value
10
10
12

Example 1: No Exceptional or Natural Events (EDT = 0)

Average = (10 + 10 + 12) / 3 = 7.333

Exceptional Data Types (EDT)

Value	EDT Code	EDT Type	EPA Concurrence
10			
10	H	EX	
12			

Example 2: Exceptional Events with No Natural Events

EDT = 1 (Exclude Events) Average = (10 + 12) / 2 = 11
EDT = 2 (Include Events) Average = (10 + 10+ 12) / 3 = 7.333
EDT = 3 (Exclude Exceptional Events) Average = (10 + 12) / 2 = 11
EDT = 4 (Exclude Natural Events) Average = (10 + 10 + 12) / 3 = 7.333
EDT = 5 (Ex. Events w/ EPA Concurrence) Average = (10 + 10 + 12) / 3 = 7.333
EDT = 6 (Ex. Exceptional w/EPA Concurrence) Average = (10 + 10 + 12) / 3 = 7.333
EDT = 7 (Ex. Natural w/EPA Concurrence) Average = (10 + 10 + 12) / 3 = 7.333

Exceptional Data Types (EDT)

Value	EDT Code	EDT Type	EPA Concurrence
10			
10	H	EX	
12	A	NAT	Y

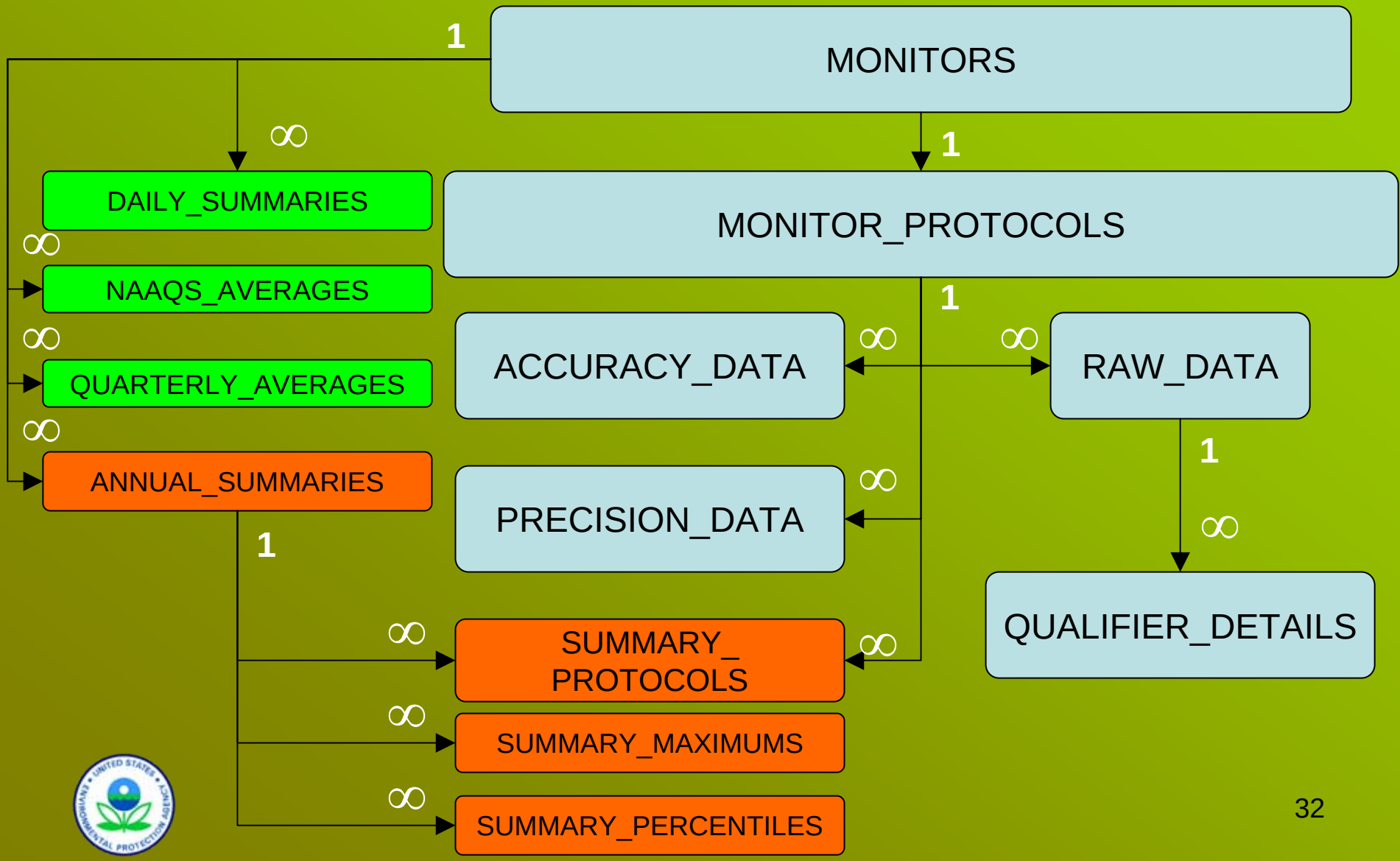
Example 3: Exceptional and Natural Events

EDT = 1 (Exclude Events) Average = (10) / 1 = 10
EDT = 2 (Include Events) Average = (10 + 10+ 12) / 3 = 7.333
EDT = 3 (Exclude Exceptional Events) Average = (10 + 12) / 2 = 11
EDT = 4 (Exclude Natural Events) Average = (10 + 10) / 2 = 10
EDT = 5 (Ex. Events w/ EPA Concurrence) Average = (10 + 10) / 2 = 10
EDT = 6 (Ex. Exceptional w/EPA Concurrence) Average = (10 + 10 + 12) / 3 = 7.333
EDT = 7 (Ex. Natural w/EPA Concurrence) Average = (10 + 10) / 2 = 10

Annual Summaries

- Basic Statistics (Averages, Standard Deviations, Counts, etc)
- Program Specific Calculations (Estimated Days Greater than the Standard, Number of Primary Exceedances, etc.)
- Top 10 Maximum Values
- Percentiles (25, 50, 75, 90, 95, 98, 99)

Sample Data Data Model



Comments

- Comments are Free-Format Text That Can Be Entered for Many Types of Information
  - SITES
  - MONITORS
  - QUALIFIER_DETAILS

Now Put it all Together….