Using Web Discoverer with AQS
ADVANCED

Jonathan Miller & Way Poteat

USEPA – National Air Data Group

2010 AQS Conference – June 8, 2010

Colorado Springs, Colorado

Items to be Covered

- Thinking About Discoverer
  - What is it Trying to Do?
  - How Do I Make it do What I Want it to Do?

- SQL Functions
  - General SQL Functions
  - Discoverer-Specific Functions
  - User Parameters

- Working With the EUL
  - What Folder to Use
  - What Filters to Keep in Mind
    - Exceptional Events, Pollutant Standard IDs, and Discoverer
    - Fixing Old Queries

- Fan Traps
  - What are they?
  - How do I Work With / Around them?

- Where to Get Additional Help

Thinking About Discoverer

- **Question:** What is it Trying to Do?

  Answer: Build a Structured Query Language (SQL) Query
  - A SQL Query has
    - "Select" Clause (What to Show)
    - "From" Clause (Where to Find the Information
    - May Have a "Where" Clause (What to Filter)
    - May Have a "Group By" Clause (if Doing Aggregate Functions)
  - Knowing a Little SQL Can Help You a LOT!!

Hints to Building a Query in Discoverer

- **STEP 1: FORM A GOOD QUESTION**
  - The Right Question Let's You Know:
    - The Items to Select (the "SELECT")
    - The Appropriate Places to Find that Data (the "FROM")
    - The Filters that Will be Required (the "WHERE")
  - Don't Be Afraid to be Annoying With Lots of Follow-up Questions!

- **STEP 2: Make Sure You and the Requestor Understand What is Being Selected**
  - What is Meant by "Active Monitors"?
  - For Ozone, do you mean the Season or the Year?
  - Is the Data You Want Based on Daily Maximums or All the Data?
  - How do You Want to Handle Events?

What is a Function?

- A Piece of Code in the Database That Can Take in Pieces of Information and Transform it Into Something Else

- Some Functions Work Differently Depending on the Types of Data They Receive

- Functions Can Be General or Custom Built for the Application
  - Currently, the AQS EUL only Supports the "General" Types of Functions
  - Example of AQS Specific Function:
    - GET_NEXT_SCHED_DATE
      - Input: a Date and a Collection Frequency Code
      - Output: The next expected date for a sample to be taken with the Provided Collection Frequency Code Based on the NAMS Schedule.

Where Can I Use Functions in Discoverer?

- Create Calculation

- Create Condition
  - Conditions Can Be Based on
    - Items
    - Values You Supply
    - Parameters (User Supplies Values at Runtime)
    - Calculations

SQL Functions

- String Functions

- Math Functions

- Date Functions

- Discoverer-Specific Functions
  - COUNT_DISTINCT

String Functions

- ||
  - StateCode||'-'||CountyCode||'-'||SiteID = 37-001-0001

- INITCAP(String)
  - INITCAP('hello') = Hello

- INSTR(String, set [, start [, occurrence]])
  - INSTR('Find it', 'i', 1, 2) = 6

- LENGTH(String)
  - LENGTH('Find it') = 7

- LOWER(String) / UPPER(String)
  - LOWER('Jonathan Miller') = jonathan miller
  - UPPER('Jonathan Miller') = JONATHAN MILLER

- LPAD(String, Length [, 'set']) / RPAD(String, Length [, 'set'])
  - LPAD('Example', 10, '-') = ---Example
  - RPAD('Example', 10, '-') = Example---

- SUBSTR(String, Start [,count])
  - SUBSTR('Find it', 3, 4) = nd i

Math Functions

- ABS – Absolute Value

- CEIL / FLOOR – Smallest Integer Above / Below Value
  - CEIL(1.6) = 2
  - FLOOR(1.6) = 1

- POWER / SQRT
  - POWER(6,2) = 36
  - SQRT(4) = 2

- ROUND/TRUNC
  - ROUND(3.859 ,2) = 3.86
  - TRUNC(3.8, 1) = 3.8

Date Functions

- TO_CHAR
  - Allows You to Format a Date Into Any Format You Want
  - TO_CHAR(sysdate, 'DD-MON-YYYY') = 08-JUN-2010

- TO_DATE
  - Converts a String Into a Date with a Defined Format
  - TO_DATE('20100608', 'YYYYMMDD')

- Date "Math"
  - Subtract a Date From a Date Gives You the Number of Days BETWEEN the Date (Non-Inclusive)
  - A Date + a Number (x) Gives You the Date "x" Days in the Future
    - TO_CHAR(SYSDATE + 1, 'DD-MON-YYYY') = 09-JUN-2010
  - You Can TRUNC or ROUND a Date
    - Returns Midnight Time
    - Default is to the Day, but Can Be Done for a Calendar Month, Quarter, or Year as Well
    - TRUNC(sysdate) = June 8, 2010
    - TRUNC(sysdate, 'MM') = June 1, 2010
    - TRUNC(sysdate, 'YYYY') = January 1, 2010

Other Date Functions

- ADD_MONTHS – Move a date x months into the future
  - ADD_MONTHS(sysdate, 2) = August 8, 2010

- LAST_DAY – Returns the date that is the last day of the month for the supplied month
  - LAST_DAY(sysdate) = June 30, 2010

- MONTHS_BETWEEN – Number of months between 2 dates
  - MONTHS_BETWEEN (TO_DATE('30-JUN-2010', 'DD-MON-YYYY'), sysdate) = 0.7833

Discoverer Specific Functions

- COUNT_DISTINCT
  - Allows You to Count Distinct Occurrences of a Field
  - Example:
    - How Many Sites Operated a PM_{10} Sampler Last Year per State?
      - Since PM_{10} Sites can be Collocated, We Don't Want to Double-Count Those Sites.
      - We'll Cover this Question in the Class Exercise
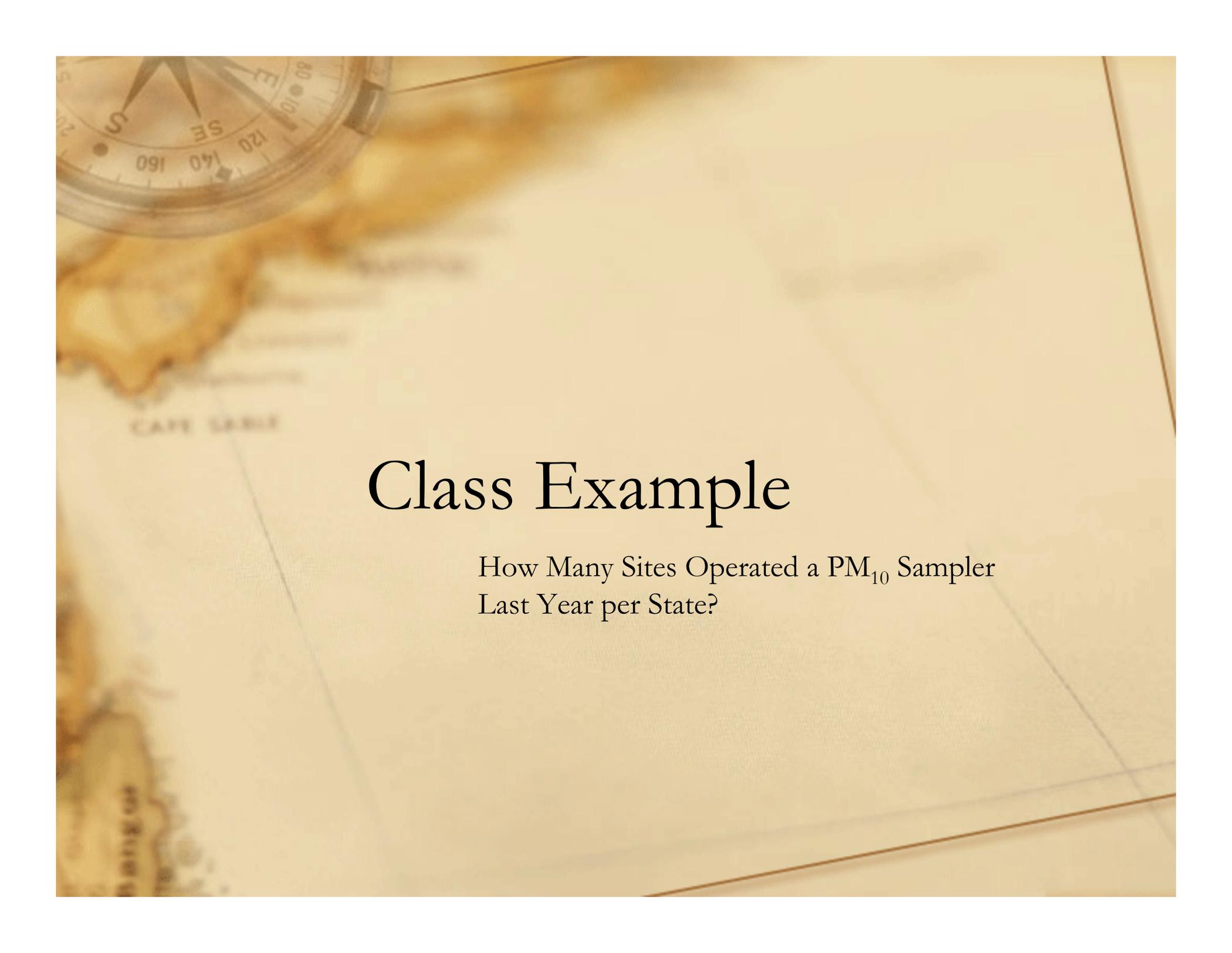
Class Example

How Many Sites Operated a PM$_{10}$ Sampler
Last Year per State?

User Parameters

- Specific Values to Use for Filters at the Time the Query is Executed

- Most Useful When Creating a Query for a Non-Discoverer User Who Will Use the Query Over and Over

The End User Layer

Working With the EUL

- The Existing EUL is Roughly Based on the AQS Data Model (One Folder Mapping to One Table in the Oracle Database)

- Know the Type of Data You Want & Where to Go to Get It

- Understand How the Data You Want Fits Into the Overall Database Structure
  - This Let's You Know How to Get There
  - Might Be a Better Place to Go

- Understand What Elements Make a Record Unique in that Table
  - Can Cause You to Get More Records Than You Expected
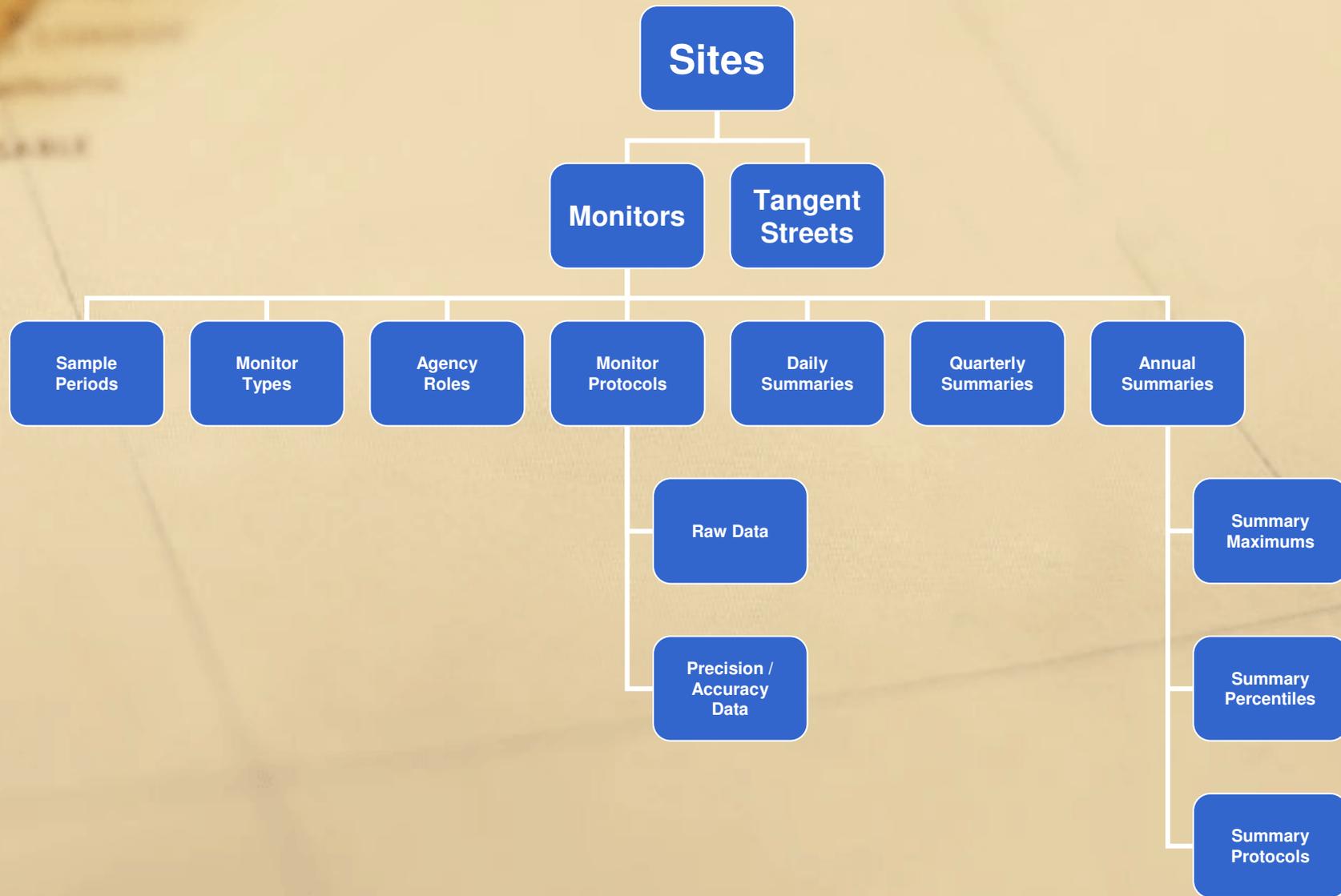
Possible Project?

- How Would You Organize the Data to be More Useful?
  - Would Pollutant Specific Folders be More Useful?
  - Thinking of Going With Fewer Folders & More Items per Folder

- Are There Types of Data That you Need to Access that You Can't Currently in Discoverer?

- Current Structure: High Flexibility / High Complexity
  - Would an EUL With Few Options be Desirable?
  - What are the Most Common Types of Things You Would Want to See?
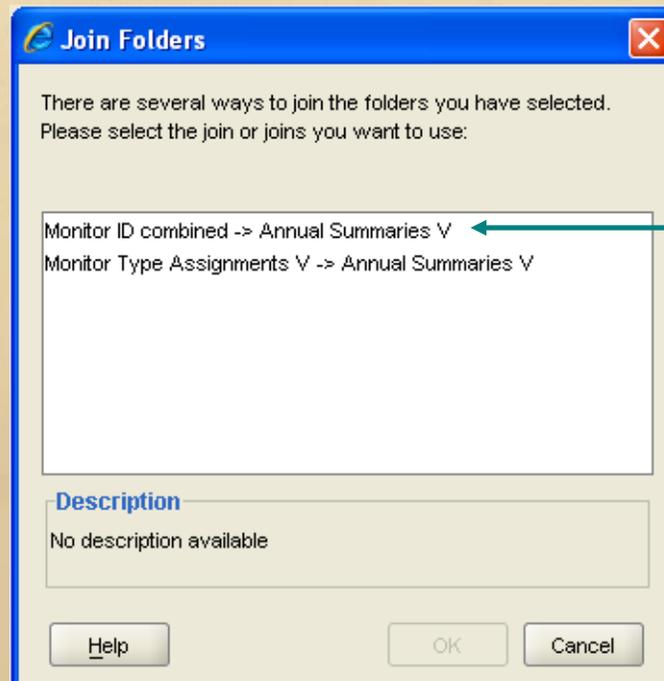
Know the Type of Data You Want

- Location Information → Sites

- Monitor Description Information → Monitors
  - Lots of Monitor Subordinate Information is Located "Under" Monitors
    - Sample Periods – When the Monitor is "Active"
    - Monitor Type Assignments – What Networks / Programs Applied
    - Agency Roles – Reporting Organizations, PQAO
    - Required Collection Frequency – How Frequently a Monitor is Required to Collect
    - Monitor Protocols – Description as to How (Instrumentation) the Data Were Collected

- Audit Information → Precision / Accuracy

- Sample Data
  - Summarized Data
    - Annual → Annual Summaries
    - Quarterly → Quarterly Summaries
    - Month (Coming Soon!!) → Monthly Summaries for Pb & PM2.5
    - Daily → Daily Summaries
  - Raw Data → Raw Data

Basic Data Structure

Sites

- **Monitors**
- **Tangent Streets**

- Sample Periods
- Monitor Types
- Agency Roles
- Monitor Protocols
  - Raw Data
  - Precision / Accuracy Data
- Daily Summaries
- Quarterly Summaries
- Annual Summaries
  - Summary Maximums
  - Summary Percentiles
  - Summary Protocols

Know How to Join

- Generally, You Want to Join by the "Parent" in the Structure

- Sometimes There are Multiple Ways to Join Between Tables. Choose the Higher "Level"

Join Folders

There are several ways to join the folders you have selected. Please select the join or joins you want to use:

Monitor ID combined -> Annual Summaries V
Monitor Type Assignments V -> Annual Summaries V

Right One

Description

No description available

Help   OK   Cancel

Filters to Keep in Mind:

- Sites
  - State + County + Site ID

- Monitors
  - *Sites* + Parameter + POC

- Summary Records
  - *Monitor ID* + Time Period + Duration + Exceptional Data Type + Pollutant Standard ID
  - Time Period Depends on the Type of Summary:
    - Annual: Year
    - Quarter: Year + Quarter
    - Month: Year + Month
    - Daily: Sample Date

- Raw Data
  - *Monitor ID* + Sample Date/Time

- P & A Data
  - *Monitor ID* + Audit Date

- Monitor Subordinate Records
  - Sample Periods
    - *Monitor ID* + Sampling Begin Date
    - Only 1 in Effect at any Given Time
  - Monitor Type Assignments
    - *Monitor ID* + Monitor Type + MT Begin Date
    - Multiple Monitor Types in Effect at a Given Time
  - Agency Roles
    - *Monitor ID* + Role + Role Begin Date
    - Only One Role per Monitor can be in Effect at a Given Time
    - Roles:
      - REPORTING
      - PQAO
      - ANALYZING
      - COLLECTING

Exceptional Events

- Two Pieces to the Exceptional Events
  - Exceptional Data Type (EDT)
    - 0: No Exceptional Events Present
    - 1: Exceptional Events Present and Excluded
    - 2: Exceptional Events Present and Included
    - 5: Exceptional Events Present and Regionally Concurred Events Excluded
  - Pollutant Standard ID
    - Events Are Associated With a Specific Standard
    - Some Pollutants Have Multiple Standards ($PM_{2.5}$ for Example)
    - Make Sure That You are Assuming the Correct Standard When Running Your Queries
    - Not all Standards Apply to All Pollutant – Duration Combinations

Pollutant Standards

- Pollutant Standard (PS) is Pollutant – Duration – Time Dependent

- Most Pollutants Have 1 PS Record at the Most per Duration
  - Exceptions:
    - SO2 Data
      - 24 Hour Standard (#6) – Applies to Duration 7 & X
      - Annual Standard (#7) – Applies to Duration 7 & 1
    - 8-Hour Ozone (Applies Only to Duration W Data)
      - 1997 Standard (#10)
      - 2008 Standard (#11)
    - $PM_{2.5}$ (Can Apply to Duration 7 or Duration X)
      - 24 hour Standard (#16)
      - Annual Standard (#18)

Fixing Old Queries

- If You Have Queries That Reference Summary Data (Any Level), You May Need to Account for the Pollutant Standard ID or You Could End Up With Multiple Rows (One Instance for Each Standard)

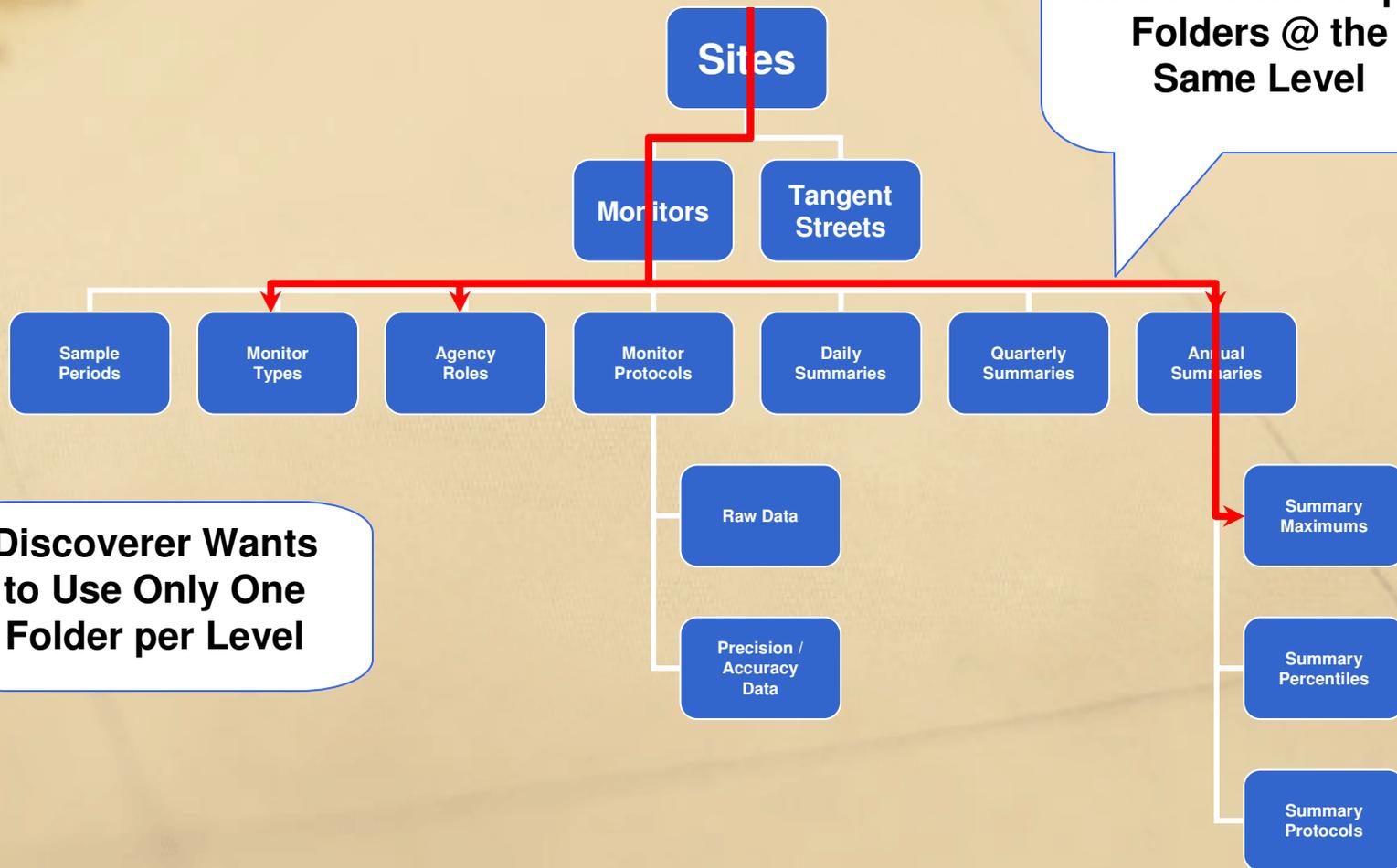- Create a Condition Where PollutantStandardID = {Appropriate Standard ID Number}

Fan Traps

- What is a "Fan Trap"?
  - Your Query has Multiple "Many-to-Many" Relationships

- How Do I Work With / Around Them?
  - Question: Do you Really Want to Disable This?
  - If "Yes" → Disable the Option in Discoverer
    - Tools → Options → "Advanced" Tab→ Disable Fan Trap Detection
  - Set Your Filters On Items to Make Sure the Right Records Are Associated Together
    - This Usually Means That You Need to Adjust the Date Ranges to Fall Within Each Other
    - Could Mean Additional Filtering Based on the Type of Information Requested

Example

- Show the Following Items
    - AQS Site ID
    - POC
    - The Monitor Type(s)
    - The PQAO in Effect
    - The Year of Data
    - The Annual Arithmetic Mean

- Filter By
    - Years Between 2008 & 2010
    - Pollutant = NO_2
    - State = Colorado

Back to the Structure:

Fan Trap Example – Selected Items

Results Without Resolving Fan Traps

Listing of Records Without the "Resolve" Filters

StateCountySiteID	POC	Summary Year (Annual)	Arithmetic Mean (Annual)	PQAO	Monitor Type
080013001	1	2008	.01734	0240	NAMS
080013001	1	2008	.01734	0240	SLAMS
080013001	1	2008	.01734	0240	UNKNOWN
080013001	1	2009	.01522	0240	NAMS
080013001	1	2009	.01522	0240	SLAMS
080013001	1	2009	.01522	0240	UNKNOWN
080013001	1	2010	.02549	0240	NAMS
080013001	1	2010	.02549	0240	SLAMS
080013001	1	2010	.02549	0240	UNKNOWN
080310002	1	2008	.02066	0240	NAMS
080310002	1	2008	.02066	0240	OTHER
080310002	1	2008	.02066	0240	SLAMS
080310002	1	2009	.03085	0240	NAMS
080310002	1	2009	.03085	0240	OTHER
080310002	1	2009	.03085	0240	SLAMS
080310002	1	2010	.03973	0240	NAMS
080310002	1	2010	.03973	0240	OTHER
080310002	1	2010	.03973	0240	SLAMS
080671004	1	2008	.00213	1110	NON-EPA FEDERAL
080671004	1	2009	.00230	1110	NON-EPA FEDERAL
080671004	1	2010	.00341	1110	NON-EPA FEDERAL
080677001	1	2008	.00328	750	TRIBAL MONITORS
080677001	1	2009	.00550	750	TRIBAL MONITORS
080677001	1	2010	.00852	750	TRIBAL MONITORS
080677003	1	2008	.00294	750	TRIBAL MONITORS
080677003	1	2009	.00474	750	TRIBAL MONITORS
080677003	1	2010	.00873	750	TRIBAL MONITORS

Results After Resolving Fan Traps

Listing of Records With the "Resolve" Filters

StateCountySiteID	POC	Summary Year (Annual)	Arithmetic Mean (Annual)	PQAO	Monitor Type
080013001	1	2008	.01734	0240	SLAMS
080013001	1	2009	.01522	0240	SLAMS
080013001	1	2010	.02549	0240	SLAMS
080310002	1	2008	.02066	0240	SLAMS
080310002	1	2009	.03085	0240	SLAMS
080310002	1	2010	.03973	0240	SLAMS
080671004	1	2008	.00213	1110	NON-EPA FEDERAL
080671004	1	2009	.00230	1110	NON-EPA FEDERAL
080671004	1	2010	.00341	1110	NON-EPA FEDERAL
080677001	1	2008	.00328	750	TRIBAL MONITORS
080677001	1	2009	.00550	750	TRIBAL MONITORS
080677001	1	2010	.00852	750	TRIBAL MONITORS
080677003	1	2008	.00294	750	TRIBAL MONITORS
080677003	1	2009	.00474	750	TRIBAL MONITORS
080677003	1	2010	.00873	750	TRIBAL MONITORS

How Did the Fan Traps Get Resolved?

- Need to Make Sure the Agency Roles and the Monitor Type Assignments Are Only Applicable for the Retrieved Annual Summary

- How? Make Sure the Role / Monitor Type Was in Effect During the Year in Question by Using the "Begin" and "End" Dates

- The Fields are Dates, and the Annual Summary Year is a Number, so we Need to Convert…

How Did the Fan Trap Get Resolved?

TO_NUMBER(TO_CHAR(Agency Role Begin Date, 'YYYY'))

TO_NUMBER(TO_CHAR(NVL(Agency Role End Date, SYSDATE), 'YYYY'))

Group Exercises