

National UIC Data Base

**Quality Data for More Effective Use
GWPC – September 28, 2010**

Beth Hall, OGWDW, EPA
Trang Le, OGWDW, EPA
Kurt Hildebrandt, Reg 7 EPA

UIC Data Base Implementation

Focus on Data Quality

- EPA's National UIC DB uses a variety of tools to screen data for accuracy, completeness and programmatic usability
- One significant benefit to using the UIC DB is the feedback provided to data submitters and EPA on data quality.
 - Beth Hall is the program lead in HQ database implementation, I'll give you a brief introduction to the UIC DB
 - Trang Le is our IT expert – she will describe some of the tools used to ensure that data flowing from state data bases to EPA meets quality thresholds'
 - I will briefly describe the collaborative QA/QC process with Nat'l Data Base participants to address programmatic sufficiency of the information with the goal of being e-reporting ready
 - Kurt Hildebrandt, EPA Region 7, will discuss how we are using the data already in the DB to program to identify and resolve implementation and reporting inconsistencies.

Overview of Nat'l Data Base

- **Launched 2008**
- **Currently Populated with Data from 20 Programs (of 69)**
 - TXRC newest with > 100 K wells
 - 270,000 wells
 - 150,000 inventory (grant eligible) wells (22 %)
 - Newest submitters include MA, SC, Navajo, EPA Region 1 and 10
 - Coming soon ! IL, OK, NE, EPA Region 5 and 9 (full payload)
- **3 Programs in e-reporting**
- **2 Programs e-reporting “ready”**
- **13 new programs initiating data mapping (6 with EN grants)**
- **Database recently upgraded (schema v2)**
 - Bigger payloads
 - More rigorous front end quality checks
 - Short term headache for submitters – EPA has conversion tool

Purpose of Nat'1 UIC Data Base

- Flow subset of well specific data from source databases to EPA
- EPA Regions and HQ need to have well specific data to manage and communicate program implementation and needs
- **How many Class II wells are Disposal (2D) vs Recovery (2R) ?**

EPA Class II Summary Data

Total

150,452

2D vs 2R

????????????

Class II Wells in Data Base (so far !) (9/22/10 – pre prod)

2D	Produced Fluid Disposal	13446
2H	Hydrocarbon Storage	643
2R	Enhanced Recovery	46941
2X	Other	1843

Trang Le

- UIC Data Base IT and Program

QA/QC of Quarterly Submission

- Submission due 45 days after end of quarter
 - Example: 3Q data through June 30 due Aug 15
 - Programs need time to update and QA program own base
- Feedback from data submitters on quality of output data is critical
 - Does information in EPA system accurately represent your source data and program?

QA/QC of Quarterly Submission

Step 1: Extract new data set

Step 2: Run standard QA/QC report

Standard set of queries

Inventory and equivalent of paper 7520 info

Step 3: Is the data complete? Is the data accurate ?

Compare to previous submission

Compare to available paper report

Ad hoc queries and analysis

How many wells have status of “proposed ?”

Do well types make sense ?

Step 5: Send data extract and QA report to the submitting program

Step 6: Hold QA/QC discussion with state, region and HQ

Reason for discrepancy

Plan to resolve

QA/QC Report

QA/QC Worksheet 3Q FY 2010 Submission

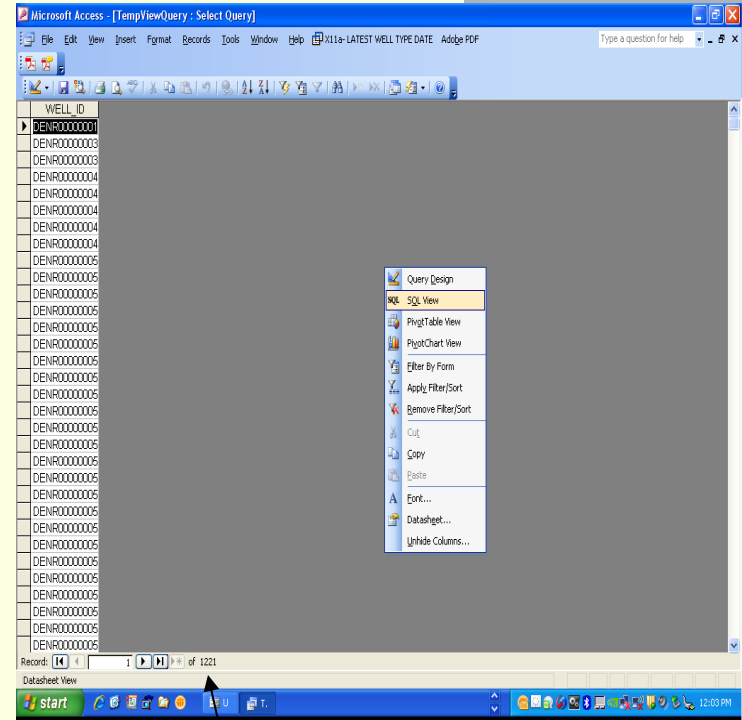
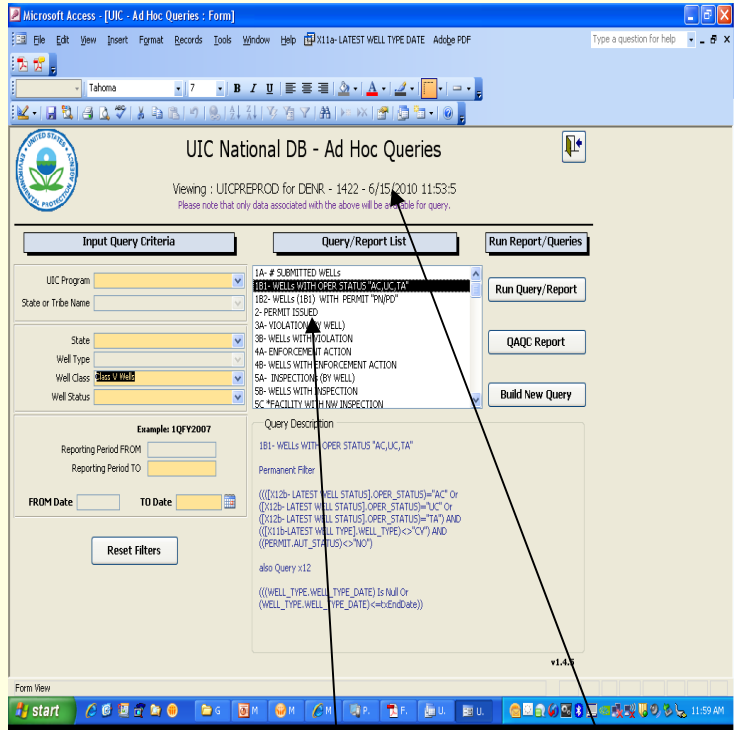
State/DI Program: _____
 Date of Query: 09/27/2010
 Date of Interview: _____
 List of Interview Participants: _____

Purpose: To access data submission quality and agree on next steps

I. Evaluate the status of well inventory

Queries	# Reported in the Database	# Reported in Other Ways (web page, paper, etc)	Issues	Resolution
1.A. How many wells in ODS- current submission	1,595			
B. How many wells in ODS = AC, TA, UC: a. THRU 12/31/2009 b. THRU Submission date?	1,202 1,247	1254 (inventory)	4 wells since 3Q	Remediation wells still not in data base
C. By well class as 12/31/2009: a. Class 1 b. Class 2 c. Class 3 d. Class 4 e. Class 5 f. (CV)	0 0 0 0 1,202 0		Many facility addresses are unknown Several wells have same lat/long	Source data being cleaned up with new inspections
D. By states/tribes as 12/31/2009				
E. Look at DATES of well type and operating status				

Demo – Semi Hypothetical



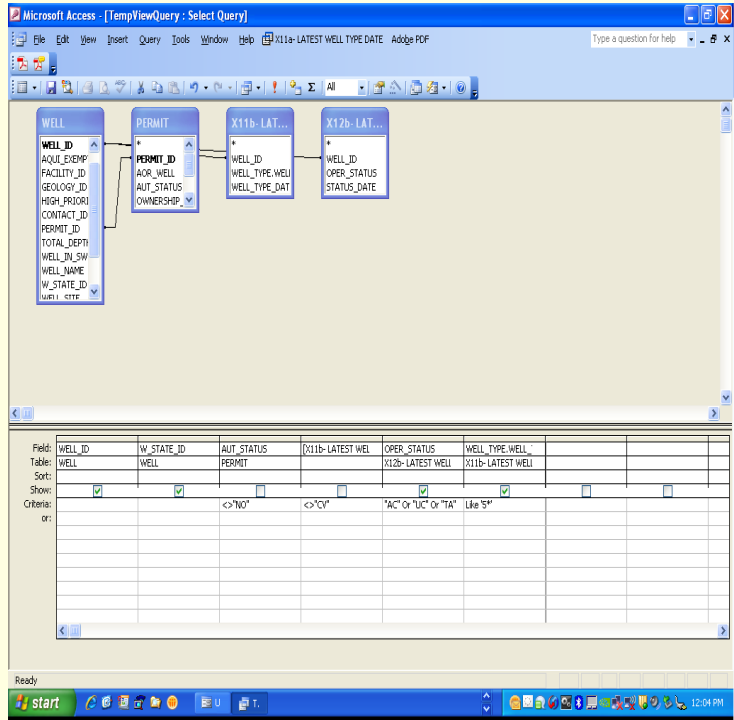
Grant eligible inventory

Extract of state data

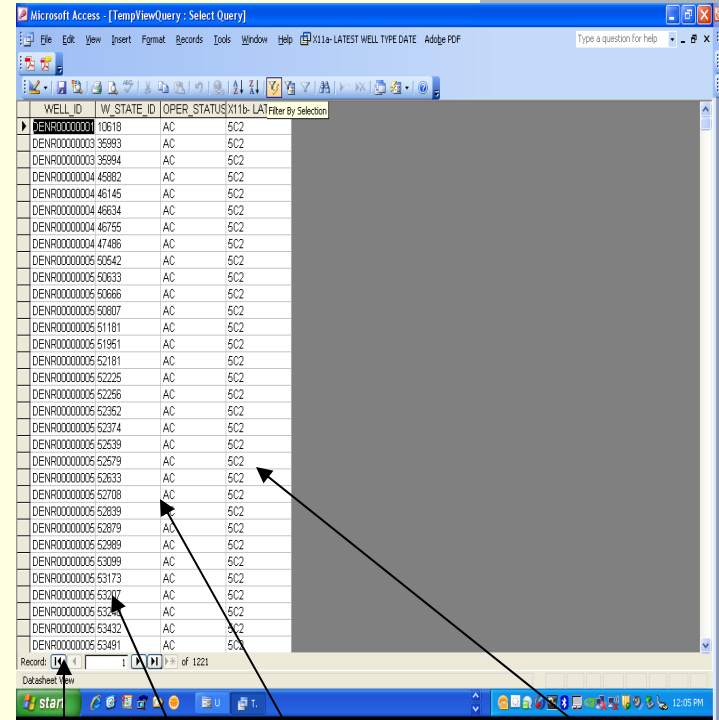
1221 Wells

“official” inventory is 1421 (hypothetical)
Why 200 well difference ?

Demo



Query logic with state well ID added



State ID
EPA ID
Well Subtype
Operating Status

Demo

AC	UC	Grand Total
5B4		3
5B6		107
5C2		956
	5D	155
	Grand Total	1221

AC	UC	Grand Total
5B4		3
5B6		107
5C2		956
	5D	155
	Grand Total	1221

State tells EPA remediation wells (5B6) accounts for 200 well difference in reported inventory and DB – Remediation wells not fully populated in state data base – should be correct within 4 months

Violations

Source data ready for e-reporting

FY 2010

Count of VIO_TYPE	
VIO_TYPE	Total
MO	14
OT	1
Grand Total	15

FY 2009

Count of WELL_ID	
VIO_TYPE	Total
MO	21
MR	4
OT	13
UI	3
Grand Total	41

FY2009 7420-2 31 (program says was incorrect)

Source data Improvement underway

FY 2010

W_STATE_ID	VIO_DATE	VIO_TYPE
239024	29-Jun-10	OT
239213	01-Mar-10	OT
239236	11-Nov-09	OT
239241	20-Jun-10	OT
239323	03-Jun-10	OT
314590	22-Jan-10	OT
239309	22-Jan-10	OT
313855	22-Jan-10	OT

FY 2009

W_STATE_ID	VIO_DATE	VIO_TYPE
239006	06-May-09	OT
239006	30-Jun-09	OT
239012	06-Apr-09	OT
239149	04-Apr-09	OT
239282	23-Jun-09	OT
239350	23-Jan-09	OT

FY 2009 7520-2 42 (MR, OM)

E-Reporting

■ Submitting Program

- Agrees UIC DB data adequately represents state database
- Agrees to flow quarterly through CDX
- Continues to update and improve source data
- Confirms data to be submitted
 - No CBI
 - Example: Well status/type changes tracked as of Jan. 2010
 - Example: No PAMs reporting

■ EPA

- Limited QA/QC check quarterly
- Produce and archive required reports (as agreed)
 - 7520's
 - Inventory
- Ad hoc analyses for program management / info requests / national trends
- Will not modify state data

Program Improvements Using the Data

- Kurt Hildebrandt
 - **Region 7 Data Management Coordinator**
 - **Region 7 UIC Program Manager**
 - **National UIC Program Expert**
 - **Chair, State / EPA DMSC**