

# CHALLENGES IN COMBINING WATER-QUALITY DATA FROM MULTIPLE AGENCIES

Gretchen P. Oelsner<sup>1</sup>, Lori A. Sprague<sup>2</sup>, Denise M. Argue<sup>3</sup>

<sup>1</sup> USGS, NM Water Science Center, Albuquerque, NM

<sup>2</sup> USGS, National Water-Quality Program, Denver, CO

<sup>3</sup> USGS, NH/VT Water Science Center, Pembroke, NH

# DATA COMPILATION

- Resources for water-quality monitoring have decreased
- Leverage information from multiple monitoring networks to address regional and national water-quality issues
- NAWQA Project of the National Water Quality Program
  - National trends in surface-water quality
- Compile publically available data from multiple sources
  - NWIS
  - STORET
  - Federal, Regional, Tribal, State, County, Local and Volunteer
- ~25 million nutrient records considered
  - ~322,000 sites
  - ~500 agencies
- Focus on nutrients but many of the same issues apply to other parameters

# DATA HARMONIZATION

- All data values needed to be described unambiguously with consistent fundamental metadata elements
  - Name
  - Physical fraction
  - Units
  - Chemical form
  - Remark codes
  - Data entry errors
- Need to have “well-mixed” data to compare across space and time
- Trend analysis is particularly sensitive to metadata issues that increase the variability in the data because it is trying to determine whether a pattern exists in the data over time
- Extra challenge given the national-scale and number of sites and sources
  - When working locally, or with a smaller number of sites, these issues are easier to resolve



# PARAMETER NAMES

- ~970 unique parameter names in the combined data
  - Ammonia – 228, Orthophosphate – 206, Nitrite + nitrate – 172
- Unambiguous names
  - Orthophosphate, water, filtered, milligrams per liter as phosphorus
  - Nitrate plus nitrite, water, filtered, milligrams per liter as nitrogen
- Ambiguous names
  - Nutrient-nitrogen – bio-available N or Total Nitrogen?
  - Phosphate, Total phosphate, Phosphate-phosphorus
    - Orthophosphate or Total Phosphorus?
    - Total phosphorus determined by converting all forms of phosphorus to phosphate
    - Some report result as Total phosphorus or Total phosphate (lab perspective)
    - Can also be used to represent orthophosphate

## POSSIBLE SOLUTIONS

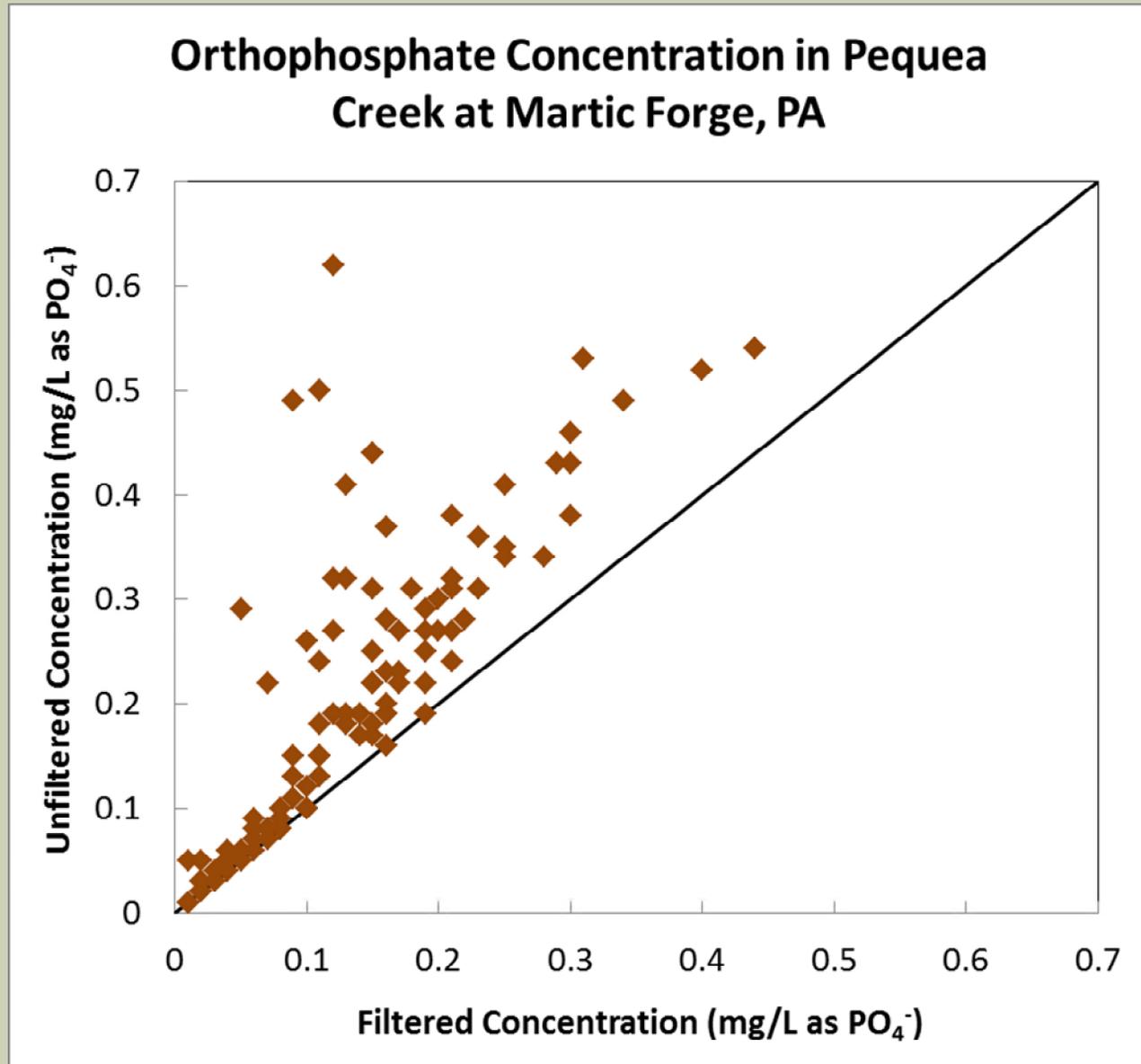
- Contact agency
- In STORET, assume Phosphate-phosphorus = Total Phosphorus

# AMBIGUOUS NAMES

| Reported Parameter Name   | Harmonized Parameter Name  |
|---|--|
| Inorganic nitrogen, water, dissolved, calculated as NH <sub>3</sub> +NO <sub>2</sub> +NO <sub>3</sub> | Inorganic Nitrogen NH <sub>3</sub> +NO <sub>2</sub> +NO <sub>3</sub> |
| Inorganic nitrogen (nitrate and nitrite)  | Nitrite + Nitrate  |
| Inorganic nitrogen (nitrate and nitrite) as N   | Nitrite + Nitrate  |
| DISSOLVED INORGANIC NITROGEN  | ?  |
| Dissolved Inorganic nitrogen as N, mg/L   | ?  |
| Inorganic N   | ?  |
| Inorganic Nitrogen  | ?  |
| NITROGEN, PARTICULATE INORGANIC   | ?  |
| Nitrogen, Inorganic   | ?  |
| Nitrogen, Inorganic     Nitrogen, inorganic, total (ug/L as N)  | ?  |
| Nitrogen, Inorganic     Nitrogen, inorganic as N  | ?  |
| Nitrogen Inorganic Total     Inorganic Nitrogen   | ?  |
| Nitrogen Inorganic Total     Total Inorganic Nitrogen, as N   | ?  |
| Total Inorganic Nitrogen  | ?  |

# PHYSICAL FRACTION

- Filtration – physical process used to separate the particulate and aqueous fractions of a water sample
- Both unfiltered and filtered variations on the same analyte can be determined for a given sample
  - Total nitrogen and Total dissolved nitrogen
- Filtered and unfiltered values for the same analyte may be very different
- Not always tied to lab method
- 44% Unambiguous fractions



*Provisional Data – Subject to Revision*

# AMBIGUOUS FRACTIONS – 56%

- Total (largest source of fraction ambiguity)
  - The inclusion of multiple species ( $\text{NH}_3$  + Organic N = total Kjeldahl nitrogen)
  - Unfiltered sample

## Unambiguous Fraction

Total Kjeldahl Nitrogen, Unfiltered

Total Kjeldahl Nitrogen, Filtered

Dissolved Total Kjeldahl Nitrogen

## Ambiguous Fraction

Total Kjeldahl Nitrogen

Kjeldahl Nitrogen

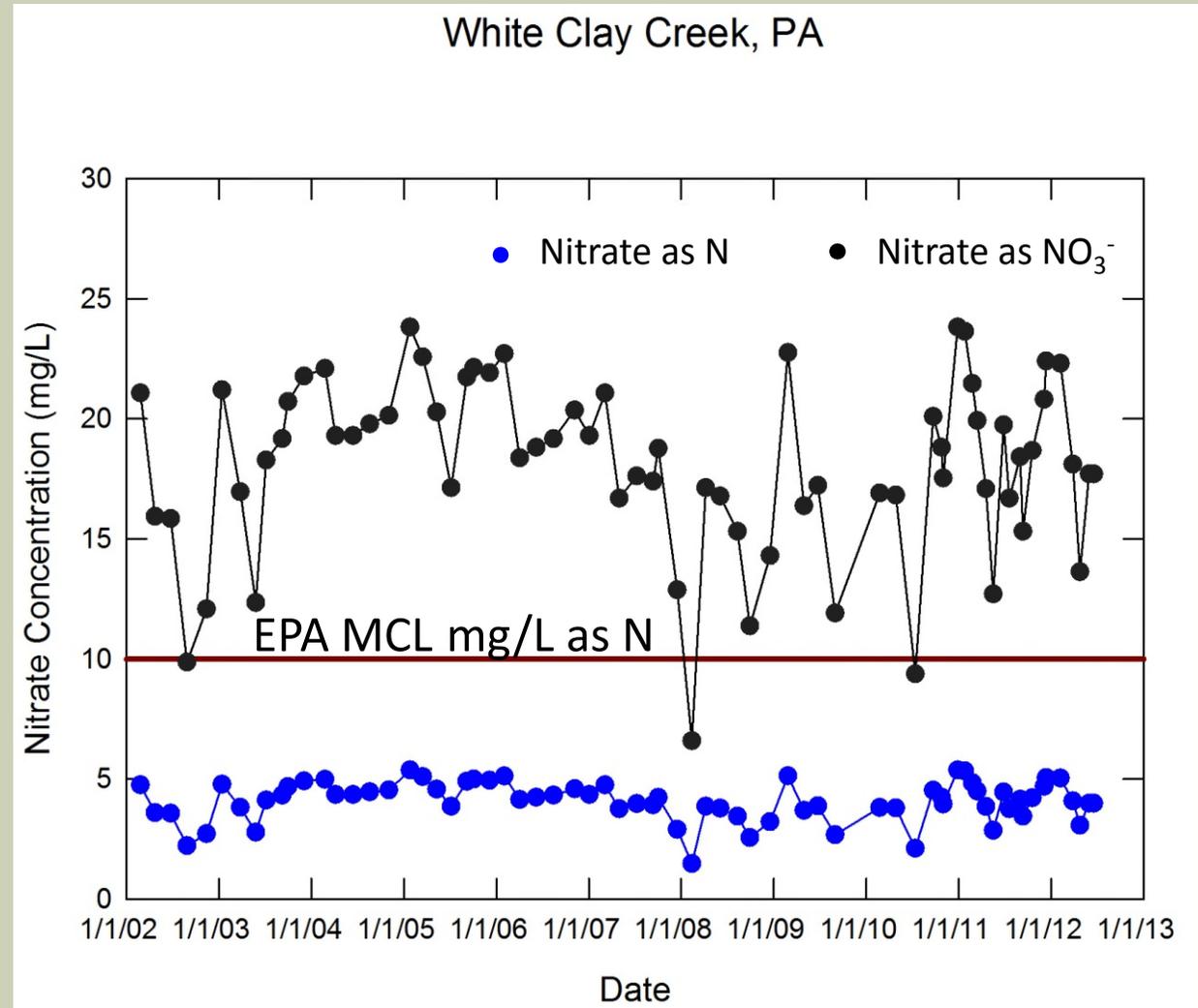
- Suspended (ambiguous) vs. particulate (unambiguous)
- Chemical fraction vs. physical fraction (acid soluble)

## POSSIBLE SOLUTIONS

- Contact agency
- Assume fraction
  - TKN, TN, TP
  - If all fractions were dissolved or total, assume total = unfiltered

# CHEMICAL FORM

- Examples:
  - as P or as P<sub>04</sub>
  - as N or as NO<sub>3</sub><sup>-</sup>
- Does it really matter?
- Can be a big difference in concentration
  - Not a function of analysis method – results can be converted and reported in many ways
- 30% of the nutrient data had incorrect/missing/ambiguous chemical form
- Contact agency



*Provisional Data – Subject to Revision*



# DATA ENTRY ERRORS

- **Missing units**
  - ~ 5% missing or incorrect units (count, CFU/100 mL, uS/cm, mg)
  - ~ 80% of the missing units were for censored (<) data
  - Do not want to bias data by not using censored data
- **Missing values**
  - Was the sample not analyzed?
  - Below detection limit?
- **Zero values**
  - Cannot have a 0 concentration
  - Often used to represent a non-detect
- **Negative values**
  - Below detection limit?
- **2.5% of the data, but 45% are for censored values**
  - Small amount of data, but potential to introduce bias

## **POSSIBLE SOLUTIONS**

- For censored data, if all non-censored values were reported in one unit, assume the same unit for censored data
- If missing + zero + negative values are <1%, replace value with the closest in time censored value

# ADDITIONAL METADATA ISSUES

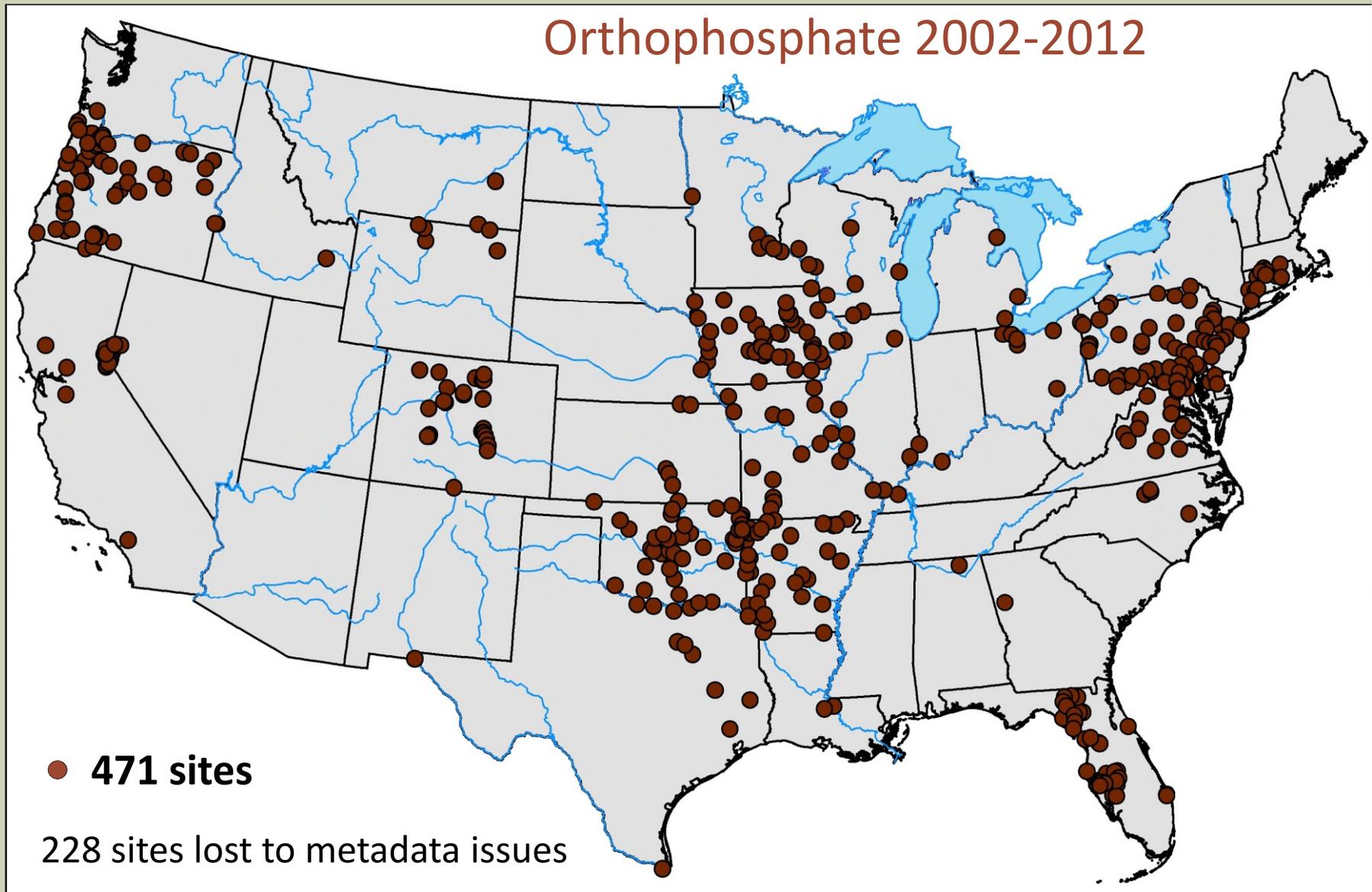
- **Outliers vs. Typos**
  - Do not want to remove natural variability or include erroneous data
  - >6 SD from the mean were assumed to be typos
- **Identifying unique sites**
  - Same agency uses different names for the same site
  - Different site names for the same physical location
- **Duplicate Data**
  - About 20% of the data are duplicates
  - Same data from 2 sources (directly from state and STORET)
  - USGS data from NWIS and another agency

**What does this mean for  
trend suitability?**

# TREND ANALYSIS SCREENING CRITERIA

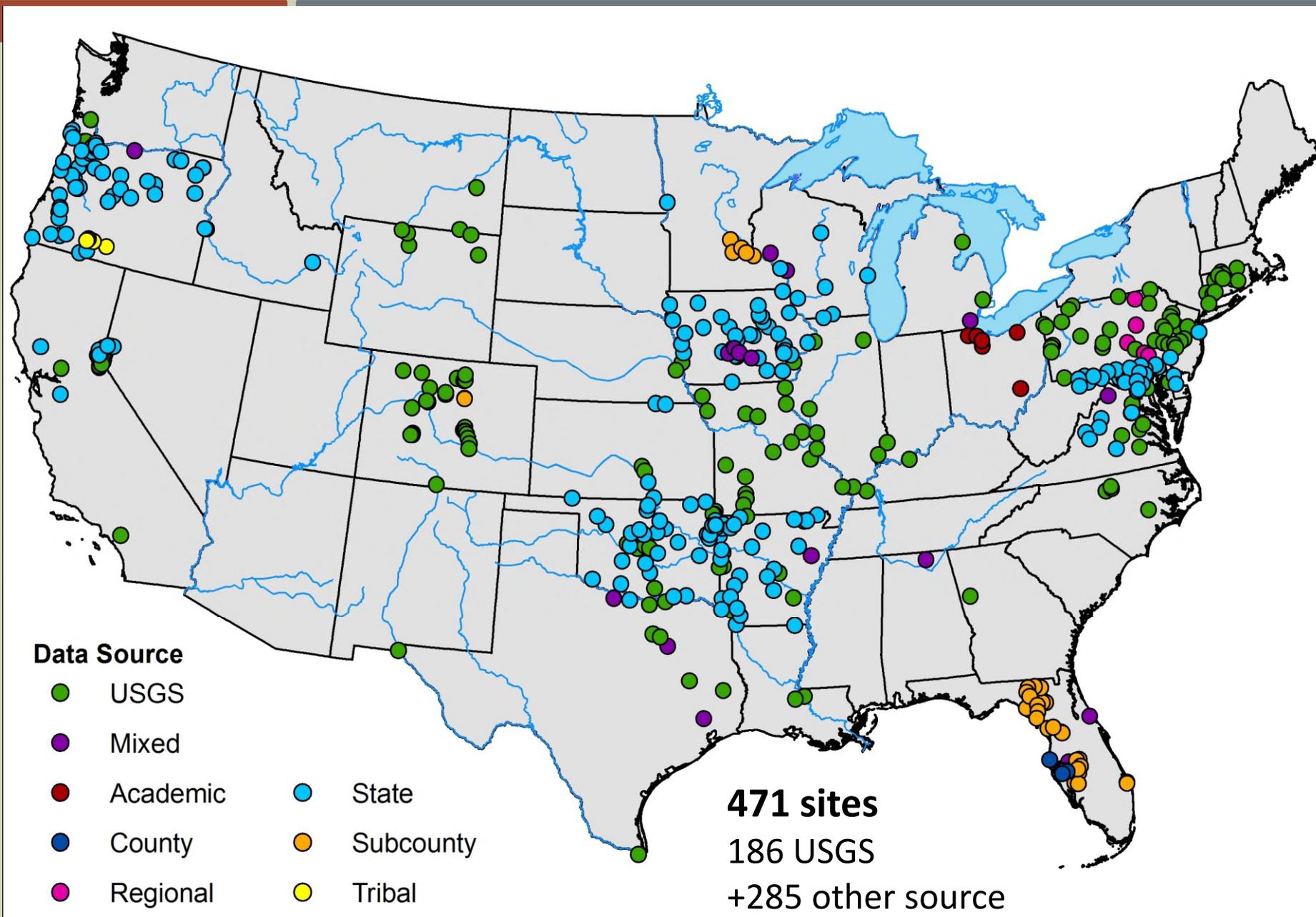
- Orthophosphate
- Trend period 2002-2012
- Seasonal samples (quarterly)
  - First 2 years
  - Last 2 years
  - 70% of the years
- Collected at or near a gage with daily discharge

# WHAT IF ALL DATA HAD METADATA?



*Provisional Data – Subject to Revision*

# ORTHOPHOSPHATE TREND SITES



# THANK YOU

- Acknowledgements

- NAWQA Surface Water Trends Team

- Lori Sprague (Coordinator/Team Lead)
- James Falcone
- Hank Johnson
- Jenny Murphy
- Gretchen Oelsner
- Karen Ryberg
- Ted Stets
- Skip Vecchia
- Bob Zuellig

- Additional Support

- Denise Argue, Melissa Riskin, Candice Hopkins, Christine Wieben

Cement Creek and Animas River Mixing



Photo: Cox, M.H., and Schemel, L.E., 2007