



Marcus W. Beck

ORISE, USEPA NHEERL Gulf Ecology Division  
Phone: 8509342480, Email: [beck.marcus@epa.gov](mailto:beck.marcus@epa.gov)

May 8, 2015

- What is NERRS/SWMP and motivation for creating the package
- What can SWMP<sub>r</sub> do
- What has SWMP<sub>r</sub> done
- How I can help, questions for the group

# What is NERRS/SWMP?

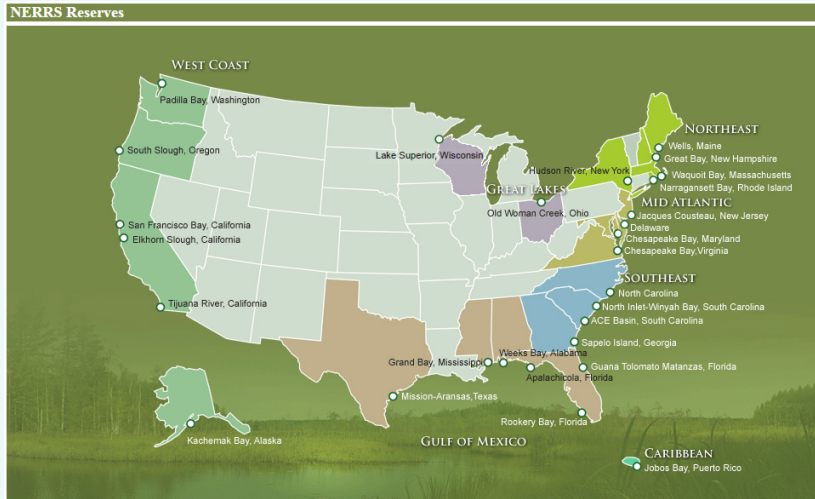
## NERRS

National Estuarine Research Reserve System, established by Coastal Zone Management Act of 1972. Focus on *long-term research*, *monitoring*, *education*, and *stewardship* for more effective coastal management.

## SWMP

System Wide Monitoring Program, initiated in 1995 to provide *continuous monitoring* data at over 140 stations in each of the 28 NERRS reserves

# What is NERRS/SWMP?



<http://nerrs.noaa.gov/ReservesMap.aspx>

# What is NERRS/SWMP?

Data maintained by the Centralized Data Management Office (CDMO)

Home	About CDMO	About Data	Get Data	Web Services	Contact CDMO
					
<b>View / Download Data</b>		<b>Real Time Monitoring Data</b>		<b>CDMO News</b>	
 <p><a href="#">Requested Citation Format</a></p>		<div>Choose Reserve... ▾</div> <p>GTMPMET 10/08/14 09:45 AM GTMPCWQ 10/08/14 09:45 AM</p>  <p>Air Temperature: 27.8 °C (82 °F) Wind Speed: 1.1 m/Sec (02 mph) Water Temperature: 22.7 °C (73 °F) Salinity: 7.1 PPT Dissolved Oxygen: 4.7 mg/L</p>		<p>The CDMO is excited to announce the launch of our new <b>SWMP Mobile application</b>. Near real-time SWMP data is now available on your smartphone or tablet at: <a href="http://www.nerrsdata.org/mobile">www.nerrsdata.org/mobile</a></p> <hr/> <p>Our <b>Data Export System</b> has been updated and now has enhanced graphing capabilities! Want to easily export or graph data? If so, check out our <a href="#">Data Export System</a></p>	

# What is NERRS/SWMP?

Each reserve has fixed, continuous monitoring stations for ***water quality*** (15 min), ***meteorology*** (15 min), and ***nutrients*** (monthly)

CDMO is an existing data management infrastructure for SWMP:

- Automated QAQC
- Numerous data download options
- Web services/API for remote data upload/retrieval
- Simple viz tools

# What is NERRS/SWMP?

As of May 1, > 58 million SWMP data records available from CDMO

Raw data will look like this...

	A	B	C	D	E	F	G	H	I	J	K	L
1	StationCo	isSWMP	DateTimeStamp	Historical	Provisional	CollMeth	REP	F_Record	PO4F	F_PO4F	NH4F	F_NH4F
2	apacpnut	P	1/10/2012 10:20	0	1	1	1		0.003	<-4> [SBL]	0.03	<0>
3	apacpnut	P	2/7/2012 11:41	0	1	1	1		0.005	<0>	0.019	<0>
4	apacpnut	P	3/5/2012 11:51	0	1	1	1		0.003	<-4> [SBL]	0.041	<0>
5	apacpnut	P	4/4/2012 10:30	0	1	1	1		0.003	<-4> [SBL]	0.043	<0>
6	apacpnut	P	5/9/2012 10:12	0	1	1	1		0.003	<0>	0.053	<0>
7	apacpnut	P	5/9/2012 10:15	0	1	1	2		0.003	<-4> [SBL]	0.022	<0>
8	apacpnut	P	5/9/2012 10:20	0	1	1	3		0.003	<0>	0.016	<0>
9	apacpnut	P	6/5/2012 8:30	0	1	1	1		0.003	<-4> [SBL]	0.04	<0>
10	apacpnut	P	7/3/2012 9:58	0	1	1	1 {CSM}		0.004	<0>	0.094	<0>
11	apacpnut	P	7/3/2012 9:59	0	1	1	2 {CSM}		0.004	<0>	0.066	<0>
12	apacpnut	P	7/3/2012 10:01	0	1	1	3 {CSM}		0.005	<0>	0.069	<0>
13	apacpnut	P	8/7/2012 9:53	0	1	1	1 {CSM}		0.003	<-4> [SBL]	0.05	<0>
14	apacpnut	P	9/5/2012 10:56	0	1	1	1		0.003	<-4> [SBL]	0.026	<0>
15	apacpnut	P	10/2/2012 9:22	0	1	1	1		0.003	<-4> [SBL]	0.042	<0>
16	apacpnut	P	10/2/2012 9:27	0	1	1	2		0.003	<-4> [SBL]	0.024	<0>
17	apacpnut	P	10/2/2012 9:32	0	1	1	3		0.003	<0>	0.042	<0>
18	apacpnut	P	11/6/2012 10:30	0	1	1	1		0.003	<-4> [SBL]	0.07	<0>
19	apacpnut	P	11/26/2012 11:39	0	1	1	1		0.003	<-4> [SBL]	0.041	<0>

# What is the problem?

An invaluable data source but no recent comparative analyses, need for tools for simple trend analysis

Some specific issues:

- Getting the data into an analysis program
- Dealing with QA/QC columns or removing ‘bad’ observations
- Combining data for comparison
- Issues inherent with time series, e.g., signal vs. noise, data quantity
- ...and analysis



# What is the (potential) solution?



An R package for SwMP data, SWMP<sub>r</sub> v2.0.0 is officially released!

```
> install.packages('SWMPr')  
> library(SWMPr)
```

Still in development, currently on v2.0.5

# SWMPPr is fully documented

## Package ‘SWMPPr’

April 1, 2015

**Type** Package

**Title** Retrieving, Organizing, and Analyzing Estuary Monitoring Data

**Version** 2.0.0

**Date** 2015-4-1

**Author** Marcus W. Beck [aut, cre]

**Maintainer** Marcus W. Beck <mbaufs2012@gmail.com>

**Description** Tools for retrieving, organizing, and analyzing environmental data from the System Wide Monitoring Program of the National Estuarine Research Reserve System. These tools address common challenges associated with continuous time series data for environmental decision making.

**BugReports** <https://github.com/fawda123/SWMPPr/issues>

**License** CC0

**Imports** data.table, http, ggmap, gridExtra, maptools, oce, dplyr, reshape2, tictoc, tidy, wq, XML

**LazyData** true

**Depends** R (>= 3.1.1), ggplot2, zoo

**NeedsCompilation** no

**Repository** CRAN

**Date/Publication** 2015-04-01 23:58:53

### R topics documented:

aggremetab	2
aggreswmp	4
all_params	5
all_params_dtrng	6
apacpnut	7
apacpwq	8
apadbwq	9
apaebmet	10

# SWMPPr is fully documented

---

qaqc	<i>QAQC filtering for SWMP data</i>
------	-------------------------------------

---

## Description

QAQC filtering for SWMP data obtained from retrieval functions, local and remote

## Usage

```
qaqc(swmpr_in, ...)  
  
## S3 method for class 'swmpr'  
qaqc(swmpr_in, qaqc_keep = 0, trace = FALSE, ...)
```

## Arguments

swmpr_in	input swmpr object
...	arguments passed to or from other methods
qaqc_keep	numeric vector of qaqc flags to keep, default 0
trace	logical for progress output on console, default FALSE

## Details

The qaqc function is a simple screen to retain values from the data with specified QAQC flags, described online: <http://cdmo.baruch.sc.edu/data/qaqc.cfm>. Each parameter in the swmpr data typically has a corresponding QAQC column of the same name with the added prefix 'f\_'. Values in the QAQC column specify a flag from -5 to 5. Generally, only data with the '0' QAQC flag should be used, which is the default option for the function. Data that do not satisfy QAQC criteria are converted to NA values. Additionally, simple filters are used to remove obviously bad values, e.g., wind speed values less than zero or pH values greater than 12. Erroneous data entered as -99 are also removed. Processed data will have QAQC columns removed, in addition to removal of values in the actual parameter columns that do not meet the criteria.

## Value

Returns a swmpr object with NA values for records that did not match qaqc\_keep. QAQC columns are also removed.

# What can SWMP<sub>r</sub> do?

SWMP<sub>r</sub> functions are grouped into three categories that describe their use in the ‘data workflow’



- Retrieve metadata
- Import from CDMO into R
- Manipulate data for analysis
- Functions to clean, combine, change time step, etc.
- Generic to specific applications
- Visualization and graphics

# What can SWMP<sub>r</sub> do?

Function types are searchable in R:

```
> help.search('analyze', package = 'SWMPr')
```

## Search Results

### Help pages:

<a href="#">SWMP<sub>r</sub>::aggremetab</a>	Aggregate metabolism data
<a href="#">SWMP<sub>r</sub>::aggreswmp</a>	Aggregate swmpr data
<a href="#">SWMP<sub>r</sub>::decomp</a>	Simple trend decomposition of swmpr data
<a href="#">SWMP<sub>r</sub>::decomp_cj</a>	Simple trend decomposition of monthly swmpr data
<a href="#">SWMP<sub>r</sub>::ecometab</a>	Ecosystem metabolism
<a href="#">SWMP<sub>r</sub>::hist.swmpr</a>	Plot swmpr using a histogram
<a href="#">SWMP<sub>r</sub>::map_reserve</a>	Map a reserve
<a href="#">SWMP<sub>r</sub>::na_approx.swmpr</a>	Linearly interpolate gaps
<a href="#">SWMP<sub>r</sub>::lines.swmpr</a>	Plot swmpr data
<a href="#">SWMP<sub>r</sub>::plot_metab</a>	Plot ecosystem metabolism for a swmpr object
<a href="#">SWMP<sub>r</sub>::plot_summary</a>	Plot graphical summaries of SWMP data
<a href="#">SWMP<sub>r</sub>::smoother</a>	Smooth swmpr data

# How are data *retrieved*?

SWMP<sub>r</sub> functions can be used to import data into R three ways

- ➊ Import from a local path
- ➋ Retrieve SWMP data from a **third-party server**
- ➌ Call the existing CDMO **web services** to import directly

Multiple options to accommodate different types of users

Bridges the gap from CDMO to analysis software

# How are data *retrieved*?

The end result is the same - data are imported as a `swmpr` data object

```
> dat <- import_remote('kacsswq')
> class(dat)

## [1] "swmpr"          "data.frame"

> head(dat, 1)

##   datetimestamp temp spcond sal do_pct do_mgl depth cdepth
## 1    2004-01-01     2    42  26   101     12   0.7     NA
##   level clevel ph turb chlfluor
## 1    NA     NA  8    6         NA

> names(attributes(dat))

## [1] "names"          "row.names"      "class"          "station"
## [5] "parameters"     "qaqc_cols"      "date_rng"       "timezone"
## [9] "stamp_class"
```

# How are data *retrieved*?

The remaining functions have `swmpr` methods

```
> methods(class = 'swmpr')  
  
## [1] aggremetab      aggreswmp      comb           decomp  
## [5] decomp_cj       ecometab       hist           lines  
## [9] na.approx       plot           plot_metab     plot_summary  
## [13] qaqc            qaqcchk        rem_reps       setstep  
## [17] smoother       subset  
## see '?methods' for accessing help and source code
```

These are functions that were written for, and work specifically, with `swmpr` objects

`swmpr` objects can also use methods from the basic data frame class, i.e., you can exit the SWMP<sub>r</sub> workflow at any time



# How are data *organized*?

Data organization depends on the analysis needs - it is usually tedious

Example: Filter by QAQC flags

- Remove observations with a specified QAQC flag value
- Remove QAQC columns: [Link](#) to QAQC codes

-5 Outside high sensor range  
-4 Outside low sensor range  
-3 Data rejected due to QAQC  
-2 Missing data  
-1 Optional SWMP supported parameter  
0 Passed initial QAQC checks  
1 Suspect data  
2 Open - reserved for later flag  
3 Calculated data: non-vented depth/level sensor correction for changes in barometric pressure  
4 Historical data: Pre-auto QAQC  
5 Corrected data

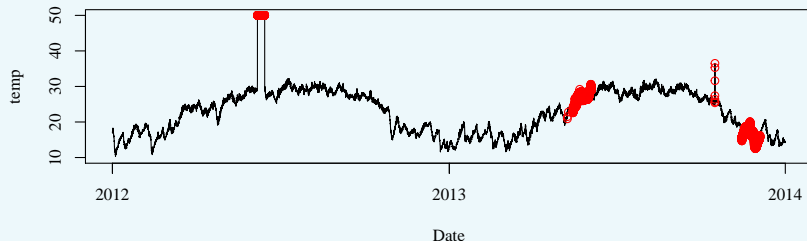
# Retrieve SWMP data

## Raw data with QAQC columns

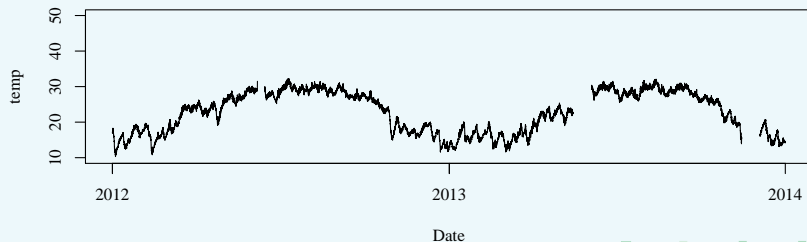
```
##          datetimestamp temp f_temp spcond f_spcond sal f_sal
## 1 2012-01-01 00:00:00   17  <0>    46    <0>   30  <0>
## 2 2012-01-01 00:15:00   17  <0>    46    <0>   30  <0>
## 3 2012-01-01 00:30:00   17  <0>    45    <0>   29  <0>
##  do_pct f_do_pct do_mgl f_do_mgl depth f_depth cdepth
## 1     89    <0>      7    <0>     1    <0>      1
## 2     88    <0>      7    <0>     1    <0>      1
## 3     89    <0>      7    <0>     1    <0>      1
##  f_cdepth level f_level clevel f_clevel ph f_ph turb
## 1    <3>    NA  <-1>    NA    NA    8 <0>      2
## 2    <3>    NA  <-1>    NA    NA    8 <0>      4
## 3    <3>    NA  <-1>    NA    NA    8 <0>      4
##  f_turb chlfluor f_chlfluor
## 1    <0>      NA    <-1>
## 2    <0>      NA    <-1>
## 3    <0>      NA    <-1>
```

# Organize SWMP data

Data in red are 'bad' QAQC flags



After using qaqc function



# How are data *organized*?

Example: Combine data to compare time series from different sites

- Data may have arbitrary time steps that do not match between sites
- Date ranges may also differ

The `comb` function addresses these issues!

```
> # import all weather and wq data for Apalachicola  
> met <- import_remote('apaebmet')  
> wq <- import_remote('apacpwq')
```

# How are data *organized*?

```
> dim(met)

## [1] 490847      11

> dim(wq)

## [1] 455808      13

> # standardize time step to two hours
> # combine only overlapping time ranges
> dat <- comb(wq, met, timestep = 120, method = 'intersect')
> dim(dat)

## [1] 56977      23
```

# How are data *organized*?

## The combined dataset

```
##          datetimestamp atemp rh    bp wspd maxwspd wdir
## 1 2001-12-31 23:00:00      4 69 1017      4      NA  347
## 2 2002-01-01 01:00:00      3 75 1017      3      NA    9
## 3 2002-01-01 03:00:00      2 77 1018      3      NA  331
## 4 2002-01-01 05:00:00      1 82 1019      4      NA    0
##    sdwdir totpar totprcp totsorad temp spcond sal do_pct
## 1      NA      0      NA      NA   NA      NA  NA      NA
## 2      NA      0      NA      NA   12     37  24    104
## 3      NA      0      NA      NA   12     40  26     99
## 4      NA      0      NA      NA   11     42  26     98
##    do_mgl depth cdepth level clevel ph turb chlfluor
## 1      NA   NA      NA   NA      NA NA   NA      NA
## 2     10     2      NA   NA      NA NA   3      NA
## 3      9     2      NA   NA      NA NA   4      NA
## 4      9     2      NA   NA      NA NA   5      NA
```

# How are data *analyzed*?

Time series analysis can range from very general to very specific

SWMP<sub>r</sub> functions include...

## *General*

- Approximate missing data
- Smoothing with moving windows
- Aggregate by time periods
- Basic plots and histograms

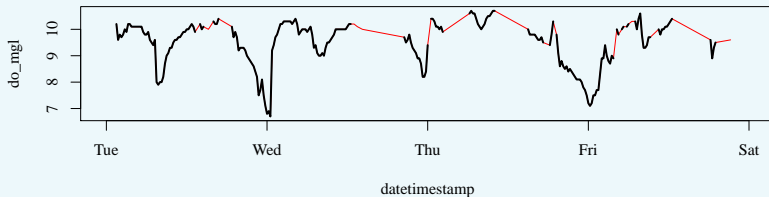
## *Specific*

- Time series decomposition
- Estimate net ecosystem metabolism
- Aggregate metabolism
- Summary plots of raw data

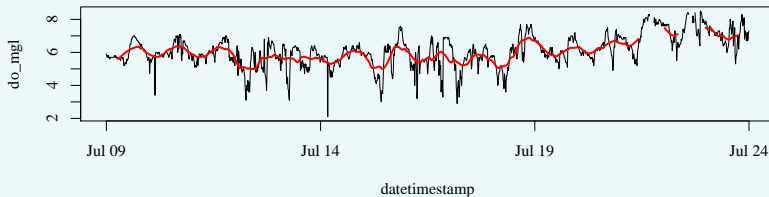
...or exit the SWMP<sub>r</sub> workflow and evaluate with other R packages

# How are data *analyzed*?

## Example: fill missing data



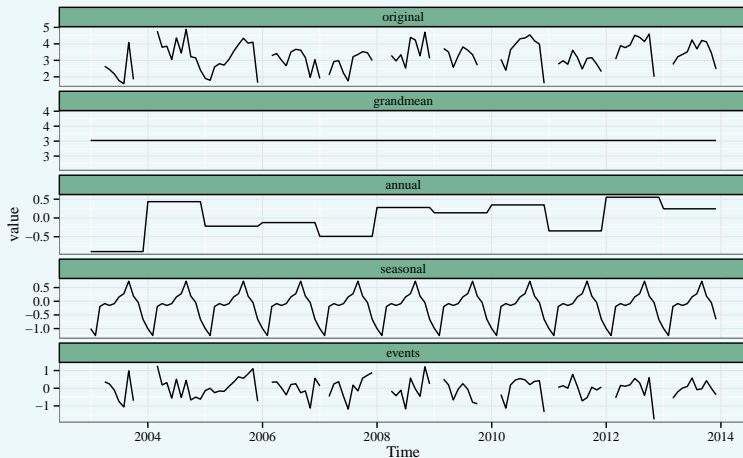
## Example: smooth data





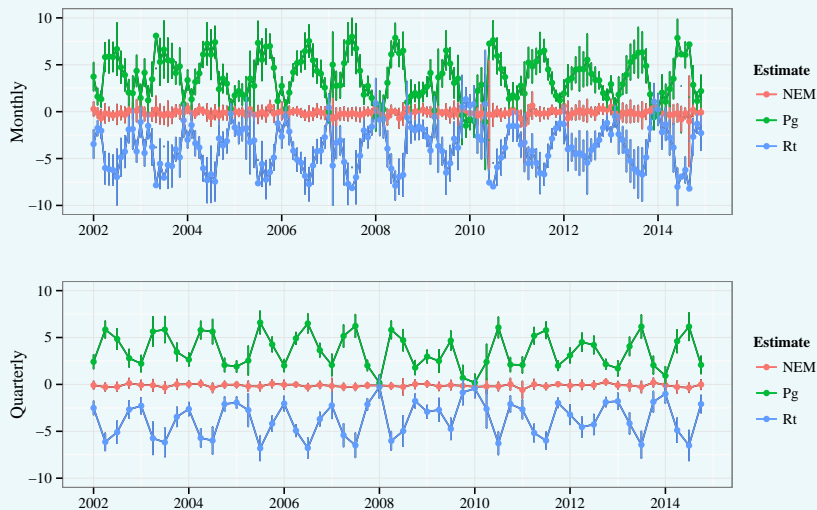
# How are data *analyzed*?

Example: time series decomposition (chl-a at cbmocnut)



# How are data *analyzed*?

## Example: estimate ecosystem metabolism



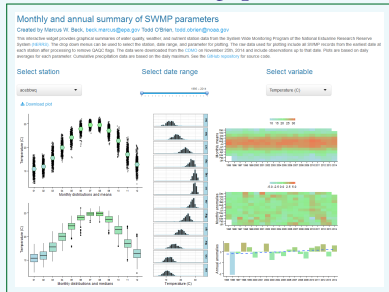
# SWMP<sub>r</sub> applications

The most common question - what is the change over time at my site?

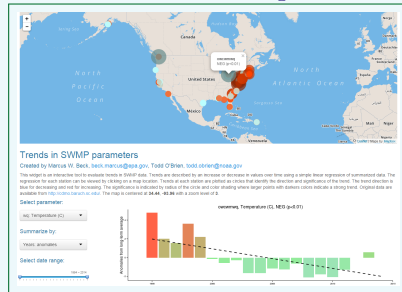
The functions in SWMP<sub>r</sub> can help, but it's easier to interact!

Two online applications can help visualize trends

## Summary plots




## Trends map



# SWMP<sub>r</sub> applications

## SWMP<sub>r</sub>ats.net: *S*ystem-*W*ide *M*onitoring *P*rogram *R*esources for the *A*nalysis of *T*ime *S*eries



**SWMP<sub>r</sub>ats.net**

- ★ **SWMP<sub>r</sub>ats.net**
- ★ **SWMP<sub>r</sub> Widgets**
  - Trends Maps
  - Summary Plots
- ★ **All about SWMP<sub>r</sub>**
- ★ **2014 Workshop**
- ★ **Forum**

**SWMP<sub>r</sub>ats.net**

The SWMP<sub>r</sub>ats.net web pages serve as a time series and data analysis information and tool resource for the National Estuarine Research Reserve System (NERRS) System-wide Monitoring Program (SWMP).



**Trends in SWMP parameters**

Created by Marcus W. Beck, [beck.marcus@epa.gov](mailto:beck.marcus@epa.gov), Todd O'Brien, [todd.obrien@noaa.gov](mailto:todd.obrien@noaa.gov)

This widget is an interactive tool to evaluate trends in SWMP data. Trends are described by an increase or decrease in values over time using a simple linear regression of summarized data. The regression for each station can be viewed by clicking on a map location. Trends at each station are plotted as circles that identify the direction and significance of the trend. The trend direction is blue for decreasing and red for increasing. The significance is indicated by radius of the circle and color shading where larger points with darker colors indicate a strong trend. Original data are available from <http://nerrs.bea.nmfs.gov> as xls. The map is centered at 38.44, -95.86 with a zoom level of 3.

# SWMPPr applications

The SWMPPr package provides an R-centric approach to *retrieve*, *organize*, and *analyze* estuary data

A new program, but already seeing heavy use:

- SWMPPr downloaded 359 times from R network (as of May 7)
- Apps used 347 hours in April

SWMPPr is meant to *augment*, not replace, existing data management programs (i.e., CDMO web services)

Deals with lots of the heavy lifting with large, unrefined datasets - greatly improves accessibility for analysis

# Novel applications

## Similarities between...

### *SWMP*

- ① Continuous wq and weather data
- ② Data from multiple sources  
- individual reserves
- ③ Interested end users -  
NOAA, NERRS  
RCs/managers

### *Stream networks*

- ① Continuous  
pressure/temperature data
- ② Data from multiple sources  
- states/tribes
- ③ Interested end users - EPA,  
states/tribes

The SWMP approach is generic enough to be adapted to other applications...

Some questions/comments...

- Data format and availability (STORET, WQX, WQP)
- Desired products - data processing vs data analysis/viz
- General questions to address - climate change...
- Distinction between data management systems vs data analysis tools, not mutually exclusive but there are key differences