

# AUTOMATED ASSESSMENT: MANAGING LARGE DATA SETS WITH R

Jack Pflaumer , Biswarup Guha, Dave Rosenman

Division of Water Monitoring & Standards  
 New Jersey Department of Environmental Protection (NJDEP)

May 29, 2018

# Overview

- Why Automation Needed?
- Manage Data
  - Data for Integrated Report
  - National Water Quality Portal (WQP)
  - Continuous Data Sets
- Why Use R?
  - New Jersey's Evolution to R
  - Benefits of R
  - Challenges and Lessons Learned





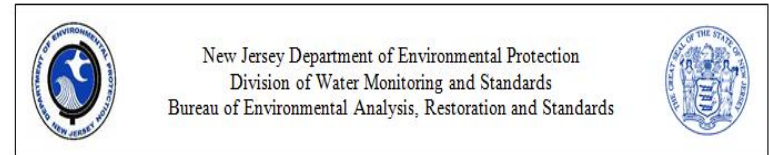
# Integrated Report Process

## ■ Chemistry Data Crunching

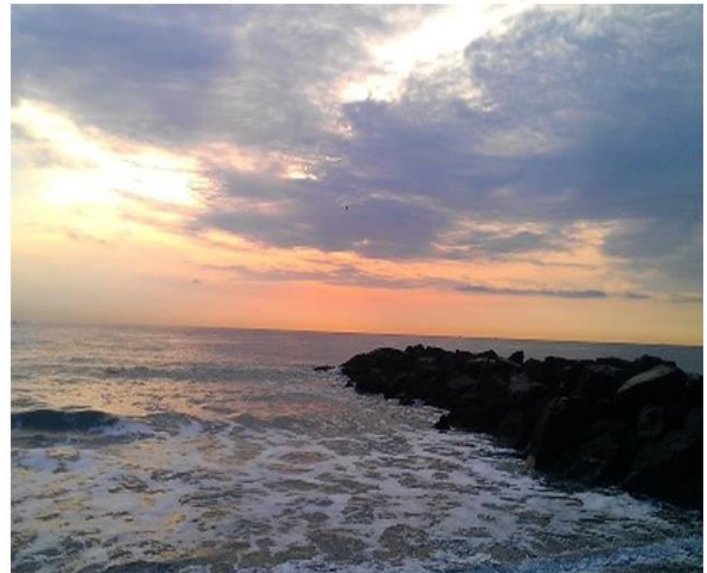
- 958 assessment units (HUC14)
- 5 - >10 years of data
- > 10,000 discrete stations
- > 3.2 million discrete data
- > 300 continuous monitoring stations
- > 90 parameters
- Biological data

## ■ Assessments

- Station level
- HUC level chemistry and biological assessment rollup
- Designated Use assessment



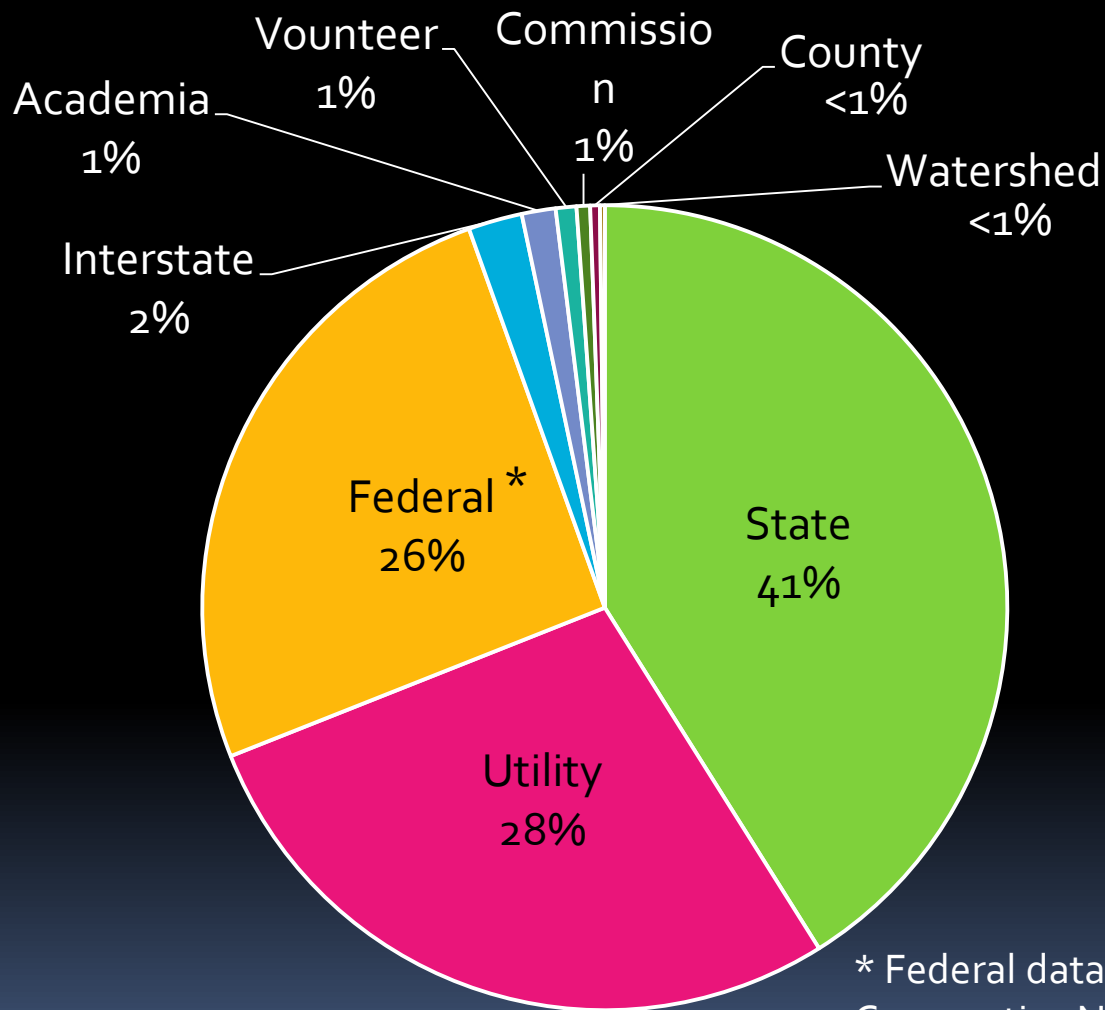
## 2014 New Jersey Integrated Water Quality Assessment Report



Atlantic Ocean at Rock Jetty, Long Branch, New Jersey  
Photo: Courtesy of Jon Dugan (AmeriCorps NJ Watershed Ambassador)

Draft  
December 2015

# 2016 NJ Integrated Report Data Sources



\* Federal data includes:  
Cooperative NJDEP/USGS  
River and Stream Monitoring Network



# Data Repositories

- NWIS
- STORET (pre-WQX)
  - Clearinghouse for nationwide water quality data
  - Difficult to import data
  - Difficult to retrieve data
  - Uneven participation and data availability
- Myriad of other databases (excel, access, text)



orgid	actid	actno	stdate	etime	local	method	medium	detcode	chainid	labfac	val	valunit
31DELRC	31DELRC	WQX	1003	Field Msr	3/22/2010	10:07:00	31DELRC_WQX	892065	USEPA	USEPA	Dissolved oxygen (DO)	8.52 mg/l
31DELRC	31DELRC	WQX	1004	Field Msr	4/20/2010	10:05:00	31DELRC_WQX	892065	USEPA	USEPA	Dissolved oxygen (DO)	9.16 mg/l
31DELRC	31DELRC	WQX	1005	Field Msr	5/25/2010	10:03:00	31DELRC_WQX	892065	USEPA	USEPA	Dissolved oxygen (DO)	9.97 mg/l
31DELRC	31DELRC	WQX	1006	Field Msr	6/22/2010	10:00:00	31DELRC_WQX	892065	USEPA	USEPA	Dissolved oxygen (DO)	6.53 mg/l
31DELRC	31DELRC	WQX	1007	Field Msr	7/20/2010	12:46:00	31DELRC_WQX	892065	USEPA	USEPA	Dissolved oxygen (DO)	6.48 mg/l
31DELRC	31DELRC	WQX	1008	Field Msr	8/30/2010	10:23:00	31DELRC_WQX	892065	USEPA	USEPA	Dissolved oxygen (DO)	7.01 mg/l
31DELRC	31DELRC	WQX	1009	Field Msr	9/27/2010	10:28:00	31DELRC_WQX	892065	USEPA	USEPA	Dissolved oxygen (DO)	6.04 mg/l
31DELRC	31DELRC	WQX	1010	Field Msr	10/26/2010	9:42:00	31DELRC_WQX	892065	USEPA	USEPA	Dissolved oxygen (DO)	8.73 mg/l
31DELRC	31DELRC	WQX	1104	Field Msr	4/26/2011	9:46:00	31DELRC_WQX	892065	USEPA	USEPA	Dissolved oxygen (DO)	10.48 mg/l
31DELRC	31DELRC	WQX	1105	Field Msr	5/24/2011	10:26:00	31DELRC_WQX	892065	USEPA	USEPA	Dissolved oxygen (DO)	8.96 mg/l
31DELRC	31DELRC	WQX	1106	Field Msr	6/21/2011	10:08:00	31DELRC_WQX	892065	USEPA	USEPA	Dissolved oxygen (DO)	6.14 mg/l
31DELRC	31DELRC	WQX	1107	Field Msr	7/19/2011	10:27:00	31DELRC_WQX	892065	USEPA	USEPA	Dissolved oxygen (DO)	6.45 mg/l
31DELRC	31DELRC	WQX	1108	Field Msr	8/22/2011	10:36:00	31DELRC_WQX	892065	USEPA	USEPA	Dissolved oxygen (DO)	3.86 mg/l
31DELRC	31DELRC	WQX	1109	Field Msr	9/14/2011	10:06:00	31DELRC_WQX	892065	USEPA	USEPA	Dissolved oxygen (DO)	7.62 mg/l
31DELRC	31DELRC	WQX	1113	Field Msr	10/17/2011	10:26:00	31DELRC_WQX	892065	USEPA	USEPA	Dissolved oxygen (DO)	9.70 mg/l
31DELRC	31DELRC	WQX	1204	Field Msr	4/29/2012	10:24:00	31DELRC_WQX	892065	USEPA	USEPA	Dissolved oxygen (DO)	8.18 mg/l
31DELRC	31DELRC	WQX	1205	Field Msr	5/22/2012	10:03:00	31DELRC_WQX	892065	USEPA	USEPA	Dissolved oxygen (DO)	7.46 mg/l
31DELRC	31DELRC	WQX	1206	Field Msr	6/25/2012	10:49:00	31DELRC_WQX	892065	USEPA	USEPA	Dissolved oxygen (DO)	6.60 mg/l
31DELRC	31DELRC	WQX	1207	Field Msr	7/23/2012	10:31:00	31DELRC_WQX	892065	USEPA	USEPA	Dissolved oxygen (DO)	4.36 mg/l
31DELRC	31DELRC	WQX	1208	Field Msr	8/21/2012	9:56:00	31DELRC_WQX	892065	USEPA	USEPA	Dissolved oxygen (DO)	4.3 mg/l
31DELRC	31DELRC	WQX	1209	Field Msr	9/24/2012	10:14:00	31DELRC_WQX	892065	USEPA	USEPA	Dissolved oxygen (DO)	6.32 mg/l
31DELRC	31DELRC	WQX	1215	Field Msr	10/22/2012	10:00:00	31DELRC_WQX	892065	USEPA	USEPA	Dissolved oxygen (DO)	7.78 mg/l
31DELRC	31DELRC	WQX	1304	Field Msr	4/23/2013	10:46:00	31DELRC_WQX	892065	USEPA	USEPA	Dissolved oxygen (DO)	10.57 mg/l
31DELRC	31DELRC	WQX	1305	Field Msr	5/21/2013	9:37:00	31DELRC_WQX	892065	USEPA	USEPA	Dissolved oxygen (DO)	6.20 mg/l
31DELRC	31DELRC	WQX	1306	Field Msr	6/25/2013	9:52:00	31DELRC_WQX	892065	USEPA	USEPA	Dissolved oxygen (DO)	7.51 mg/l
31DELRC	31DELRC	WQX	1306	Field Msr	6/25/2013	9:52:00	31DELRC_WQX	892065	USEPA	USEPA	Dissolved oxygen (DO)	7.51 mg/l

The dissolved arsenic levels at these site were as follows:

01444990  
 2/8/2005: 0.2 ug/L  
 8/16/2005: 0.6 ug/L, sediment (tot rec.) 1.1 ug/g

01445100  
 2/16/2005: 0.3 ug/L, sediment (tot rec.) 1.4 ug/g  
 8/11/2005: 0.8 ug/L,

01445700  
 2/17/2005: 0.3 ug/L  
 8/24/2005: 0.7 ug/L

All total recoverable samples were <2.0 ug/L.





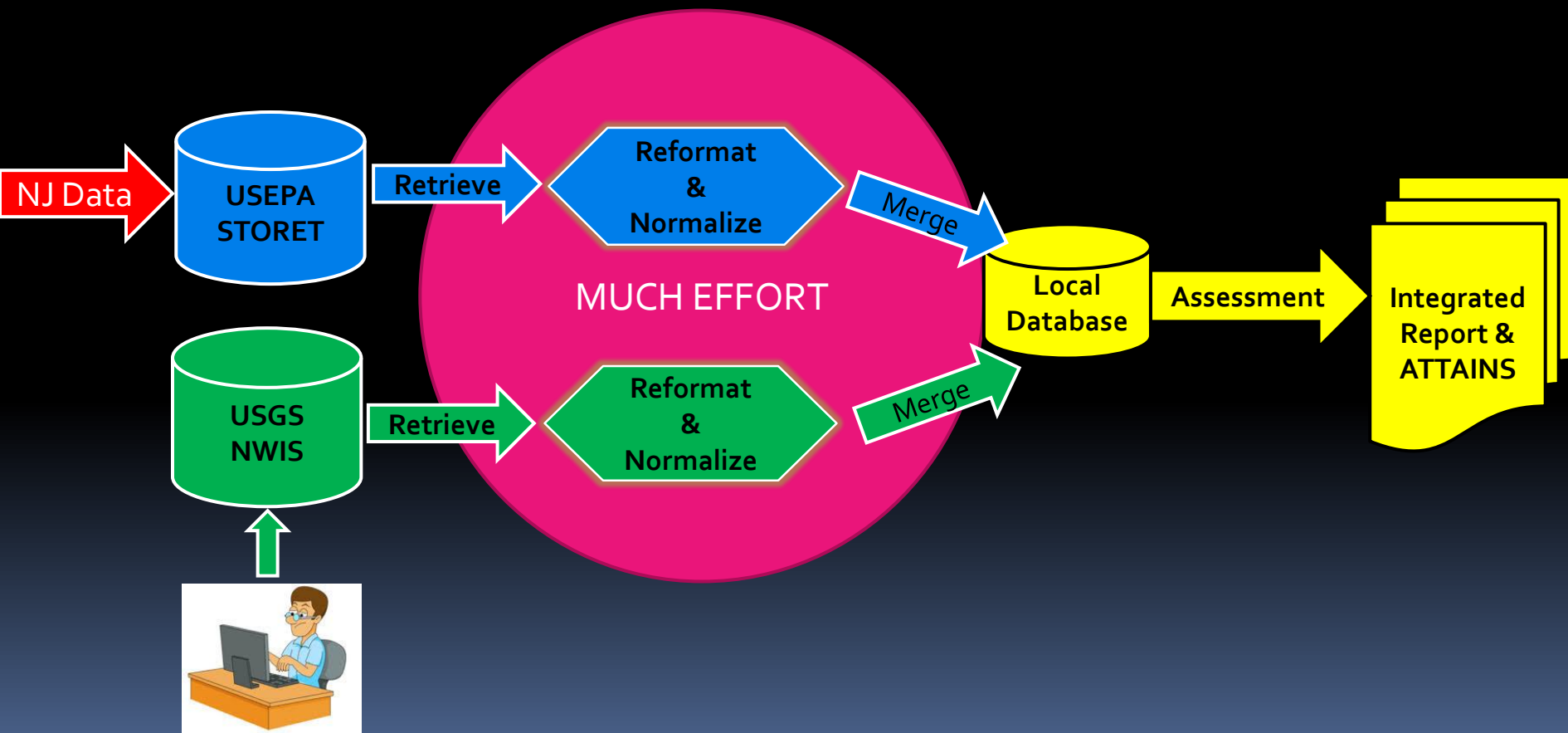
# Water Quality Portal Benefits



- Since 2012: EPA (STORET), USGS (NWIS), USDA (STEWARDS)
- One Stop Shop for data (except for continuous)
- Standardized data protocol
- Easy to use interface with help desk
- Closer collaboration with stakeholders with WQX
- Publicly available and easily accessible
- NJ requires approved QAPP

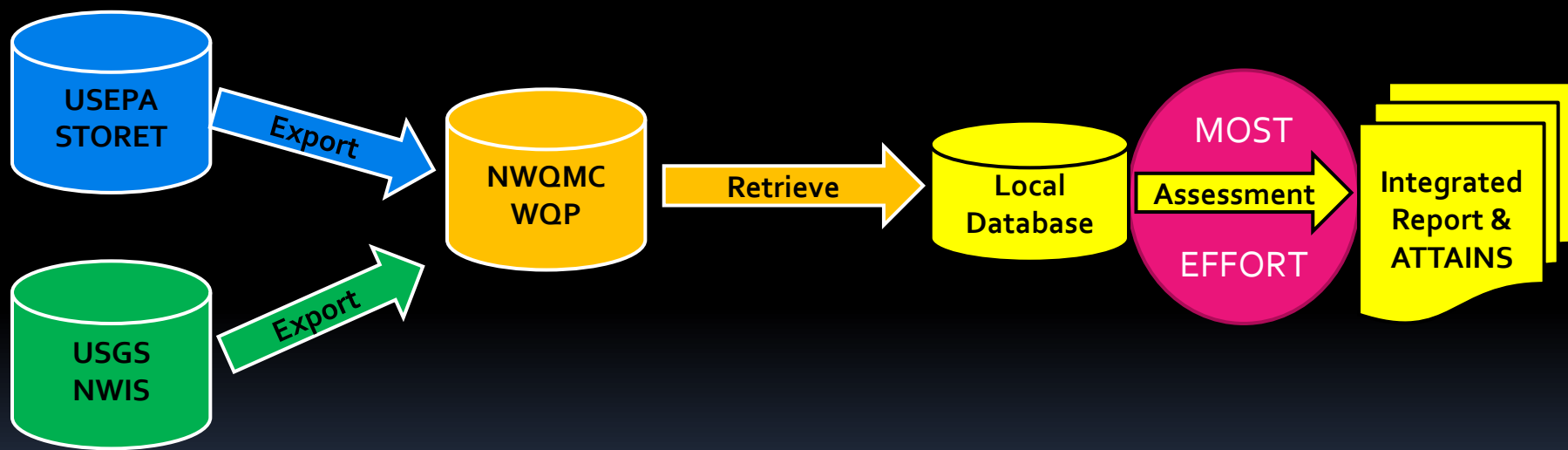


# Data Access & Assessment BEFORE the Water Quality Portal





# Data Access & Assessment in AFTER the Water Quality Portal



# Continuous Data Solution

- National Issues with Continuous Monitoring Data
  - STORET not designed for this
  - No data standard
  - Difficulty sharing data
- NJ Solution – web accessible database (Rutgers U)
  - Interactive
  - Easy to use
  - Expandability
    - Other data generators
    - EPA is working with Rutgers

Governor Chris Christie • Lt. Governor Kim Guadagno  
NJ Home | Services A to Z | Departments/Agencies | FAQ  
Search [All of NJ] [v]

STATE OF NEW JERSEY  
DEPARTMENT OF ENVIRONMENTAL PROTECTION  
DWM&S Continuous Data Monitoring Program

DEP Home | About DEP | Index by Topic | Programs/Units | DEP Online

In addition to its many discrete monitoring activities, the Division of Water Monitoring & Standards (DWM&S) also manages numerous continuous monitoring activities that involve the use of short and long term data loggers throughout NJ. In both fresh and marine waters, data loggers are electronic devices that utilize sensors and probes to measure various water quality parameters 24 hours per day, 7 days per week. Data logger deployments may range from several days to many months. Sites for continuous monitoring are chosen for a variety of reasons, including to support and/or supplement the discrete data collected by DWM&S, as well as for project-specific purposes.

Continuous monitoring data in NJ are used for a variety of purposes, including measuring fluctuations of Dissolved Oxygen and Temperature to determine the ecological health of streams, acquisition of data for comprehensive analysis and decision making to support [Barnegat Bay's Comprehensive Action Plan](#), and generation of the [New Jersey Integrated Water Quality Monitoring and Assessment Report \(Integrated Report\)](#). Every two years, a new water region in NJ is given a comprehensive monitoring and assessment focus which starts new continuous monitoring site selection.

On this website, users will find the locations of DWM&S managed continuous monitoring activities – both long and short term, active and historical, as well as information specific to each location. Data associated with each location is available for download and graphing.

Map | Data | Graph

NOTE: Provisional Data Subject to Revision.

Contact DEP | Privacy Notice | Legal Statement & Disclaimers | Accessibility Statement

Water Monitoring & Standards: Home  
Department: [RUCDP Home](#) | [About DEP](#) | [Index by Topic](#) | [Programs/Units](#) | [DEP Online](#)  
Statewide: [NJDEP Web Site](#) | [NJDEP Data Portal](#) | [NJDEP Data Hub](#)  
Copyright © State of New Jersey, 1996-2016  
Last Updated: March 10, 2015



# Continuous Data Solution

**AN0521 Short-Term Historical**  
 Description: MAPLE ROOT BR BOWMAN RD  
 Latitude: 40.08125  
 Longitude: -74.32714  
 Affiliation: NJDEP - Freshwater & Biological Monitoring  
 Affiliation URL: [www.state.nj.us/dep/wms/bfbm/index.html](http://www.state.nj.us/dep/wms/bfbm/index.html)  
**Data Date Range:**  
 Earliest: 2015-06-26 01:30:00  
 Latest: 2015-10-01 09:30:00  
[Graph Download](#)

### Station Graphing

Please select the primary station you wish to look at and a variable.

Select Data Set: QA'd Data  
 Select Station: BB04a  
 Select Variable 1: DO Concentration  
 Select Variable 2: Salinity  
 Date Range:  Entire Date Range  Latest 2 Weeks  Latest 5 Day  Latest 3 Day  Custom  
 QA Date range: 07/19/2012 to 09/08/2015  
[Graph](#)

Legend: — DO Concentration [mg/L] — Salinity [ppt]

**Choose Geography**  
 County: OCEAN  
 Municipalities: Jackson Township  
 Huc14: - All -  
 Station Type: Short-Term Historical  
 Stations: AN0521  
 Start Date: - Month - - Day - - Year - [Clear](#)  
 End Date: - Month - - Day - - Year - [Clear](#)

**Parameters:**  
 Select all parameters  
 Temperature  Specific Conductance  Salinity  DO %  
 DO Conc  DO Charge  pH  Turbidity  Chlorophyll

[Download Data](#) Records Selected: 4,673

Legend:  
 ● Long-Term Active  
 ● Long-Term Historical  
 ● Offline  
 ● Short-Term Historical

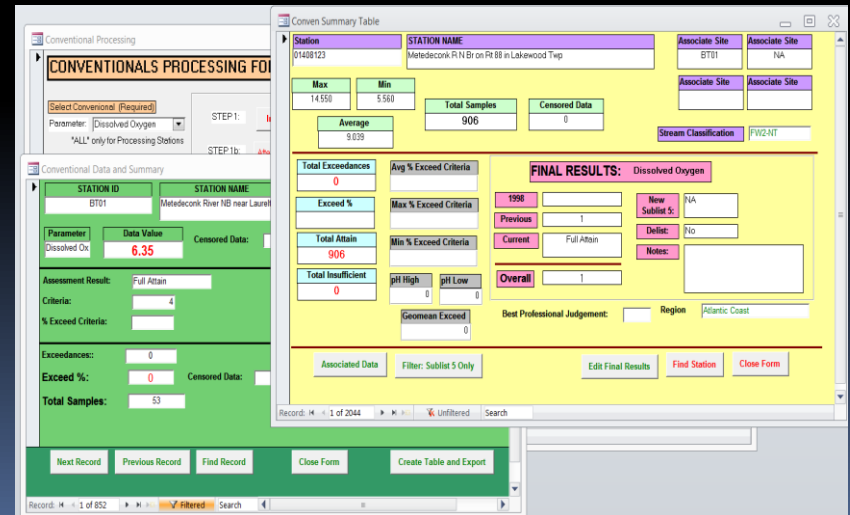
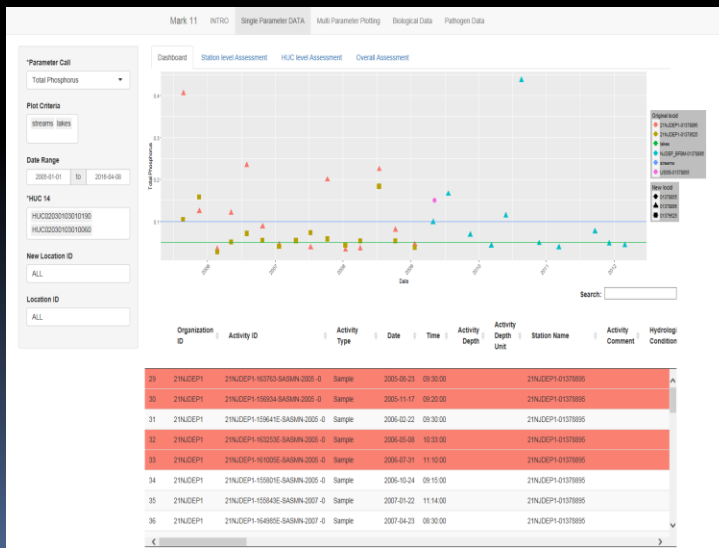
**NOTE: Provisional Data Subject to Revision.**



# History of Automated Assessment Process

- Incremental improvements to automation process
- Excel (pre-2006)
- Access (2006)
- R (2016)

A	B	C	D	E	H	I	M	N	O	P	Q	R
orgid	actid	acttyp	stdate	sttime	locid	metcont	metnam	detcont	charnam	smfrac	val	valunit
31DELRBC	31DELRBC_WQX-1003	Field Msr/l	3/22/2010	10:07:00	31DELRBC_WQX-892065	USEPA	USEPA		Dissolved oxygen (DO)		8.52	mg/l
31DELRBC	31DELRBC_WQX-1004	Field Msr/l	4/20/2010	10:05:00	31DELRBC_WQX-892065	USEPA	USEPA		Dissolved oxygen (DO)		9.16	mg/l
31DELRBC	31DELRBC_WQX-1005	Field Msr/l	5/25/2010	10:03:00	31DELRBC_WQX-892065	USEPA	USEPA		Dissolved oxygen (DO)		9.97	mg/l
31DELRBC	31DELRBC_WQX-1006	Field Msr/l	6/22/2010	0:00:00	31DELRBC_WQX-892065	USEPA	USEPA		Dissolved oxygen (DO)		6.53	mg/l
31DELRBC	31DELRBC_WQX-1007	Field Msr/l	7/20/2010	12:46:00	31DELRBC_WQX-892065	USEPA	USEPA		Dissolved oxygen (DO)		6.48	mg/l
31DELRBC	31DELRBC_WQX-1008	Field Msr/l	8/30/2010	10:23:00	31DELRBC_WQX-892065	USEPA	USEPA		Dissolved oxygen (DO)		7.01	mg/l
31DELRBC	31DELRBC_WQX-1009	Field Msr/l	9/27/2010	10:28:00	31DELRBC_WQX-892065	USEPA	USEPA		Dissolved oxygen (DO)		6.04	mg/l
31DELRBC	31DELRBC_WQX-1010	Field Msr/l	10/26/2010	9:42:00	31DELRBC_WQX-892065	USEPA	USEPA		Dissolved oxygen (DO)		8.73	mg/l
31DELRBC	31DELRBC_WQX-1104	Field Msr/l	4/26/2011	9:46:00	31DELRBC_WQX-892065	USEPA	USEPA		Dissolved oxygen (DO)		10.48	mg/l
31DELRBC	31DELRBC_WQX-1105	Field Msr/l	5/24/2011	10:26:00	31DELRBC_WQX-892065	USEPA	USEPA		Dissolved oxygen (DO)		8.96	mg/l
31DELRBC	31DELRBC_WQX-1106	Field Msr/l	6/21/2011	10:08:00	31DELRBC_WQX-892065	USEPA	USEPA		Dissolved oxygen (DO)		6.14	mg/l
31DELRBC	31DELRBC_WQX-1107	Field Msr/l	7/19/2011	10:27:00	31DELRBC_WQX-892065	USEPA	USEPA		Dissolved oxygen (DO)		6.45	mg/l
31DELRBC	31DELRBC_WQX-1108	Field Msr/l	8/21/2011	10:36:00	31DELRBC_WQX-892065	USEPA	USEPA		Dissolved oxygen (DO)		5.86	mg/l
31DELRBC	31DELRBC_WQX-1109	Field Msr/l	9/14/2011	10:06:00	31DELRBC_WQX-892065	USEPA	USEPA		Dissolved oxygen (DO)		7.62	mg/l
31DELRBC	31DELRBC_WQX-1110	Field Msr/l	10/17/2011	10:26:00	31DELRBC_WQX-892065	USEPA	USEPA		Dissolved oxygen (DO)		9.76	mg/l
31DELRBC	31DELRBC_WQX-1204	Field Msr/l	4/23/2012	10:24:00	31DELRBC_WQX-892065	USEPA	USEPA		Dissolved oxygen (DO)		8.3	mg/l
31DELRBC	31DELRBC_WQX-1205	Field Msr/l	5/22/2012	10:03:00	31DELRBC_WQX-892065	USEPA	USEPA		Dissolved oxygen (DO)		7.46	mg/l
31DELRBC	31DELRBC_WQX-1206	Field Msr/l	6/25/2012	10:49:00	31DELRBC_WQX-892065	USEPA	USEPA		Dissolved oxygen (DO)		6.65	mg/l
31DELRBC	31DELRBC_WQX-1207	Field Msr/l	7/23/2012	10:33:00	31DELRBC_WQX-892065	USEPA	USEPA		Dissolved oxygen (DO)		4.36	mg/l
31DELRBC	31DELRBC_WQX-1208	Field Msr/l	8/21/2012	9:56:00	31DELRBC_WQX-892065	USEPA	USEPA		Dissolved oxygen (DO)		4.3	mg/l
31DELRBC	31DELRBC_WQX-1209	Field Msr/l	9/24/2012	10:14:00	31DELRBC_WQX-892065	USEPA	USEPA		Dissolved oxygen (DO)		6.12	mg/l
31DELRBC	31DELRBC_WQX-1210	Field Msr/l	10/22/2012	10:00:00	31DELRBC_WQX-892065	USEPA	USEPA		Dissolved oxygen (DO)		7.78	mg/l
31DELRBC	31DELRBC_WQX-1304	Field Msr/l	4/23/2013	10:46:00	31DELRBC_WQX-892065	USEPA	USEPA		Dissolved oxygen (DO)		10.57	mg/l
31DELRBC	31DELRBC_WQX-1305	Field Msr/l	5/21/2013	9:37:00	31DELRBC_WQX-892065	USEPA	USEPA		Dissolved oxygen (DO)		6.26	mg/l
31DELRBC	31DELRBC_WQX-1306	Field Msr/l	6/25/2013	9:52:00	31DELRBC_WQX-892065	USEPA	USEPA		Dissolved oxygen (DO)		7.51	mg/l
31DELRBC	31DELRBC_WQX-1306	Field Msr/l	6/25/2013	9:52:00	31DELRBC_WQX-892065	USEPA	USEPA		Dissolved oxygen (DO)		7.51	mg/l

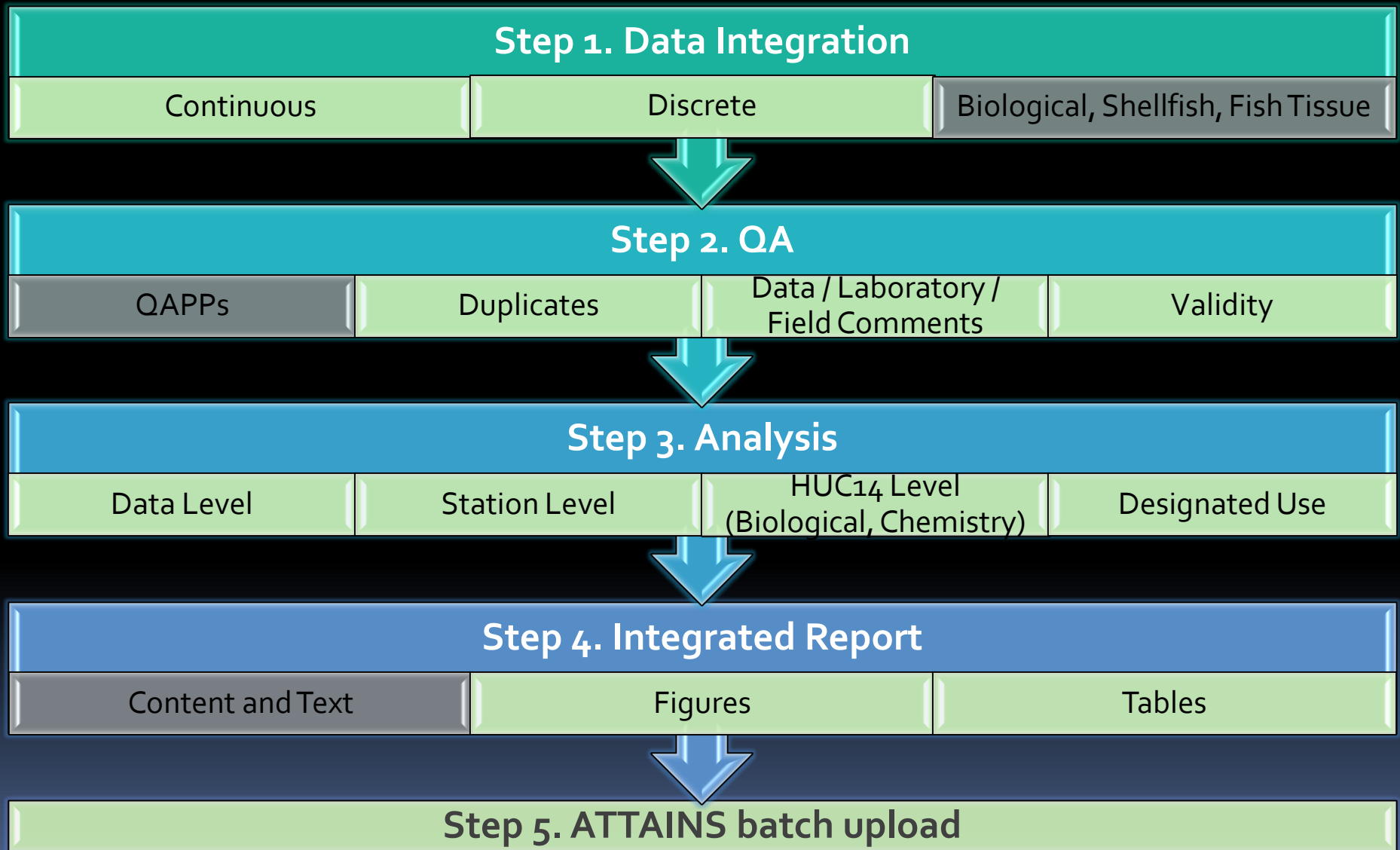


# Why Manage Data with R

- Free software
- Easily handles large datasets, easy to use
- No special data management system needed
- Store data in ordinary ascii or binary files
  - Large datasets – binary format improved efficiency
- Automated downloads (WQP, USGS)
- Vast community and resources available
- Can do everything with R: data management, assessment, reporting
- Integrates with other languages and platforms

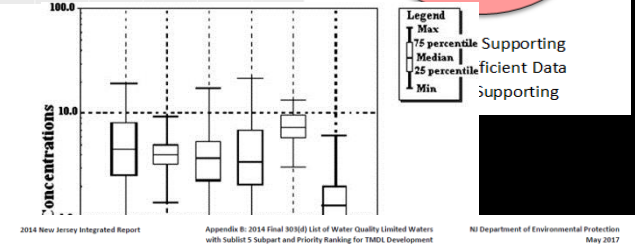
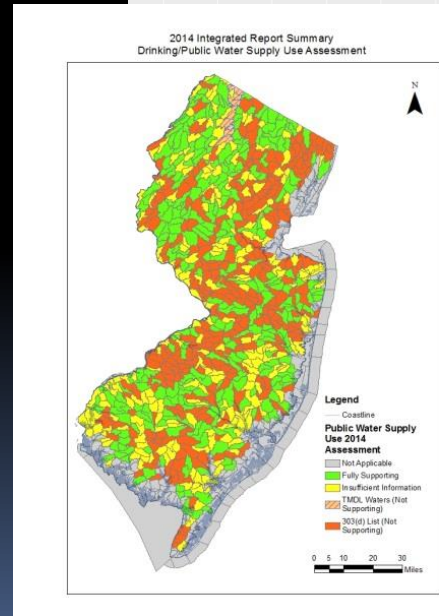
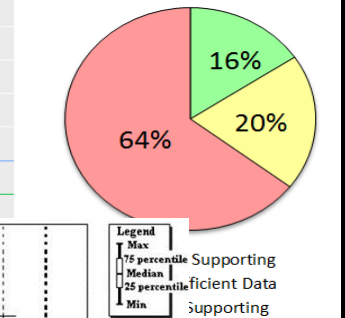
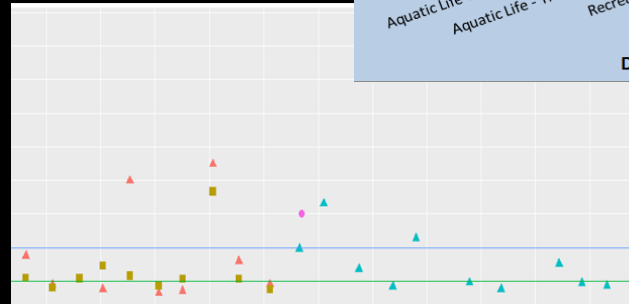
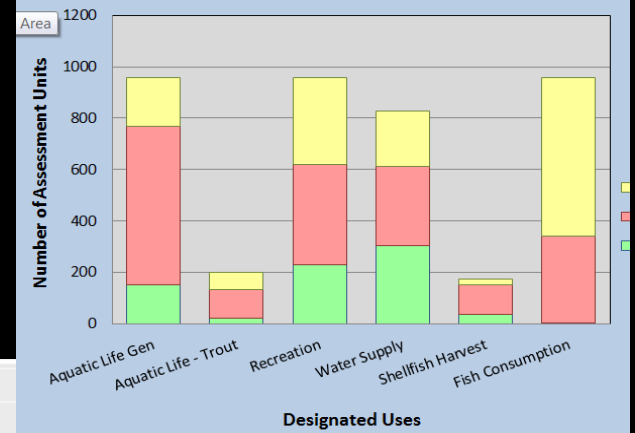


# NJ IR Automated Process



# Reporting Tools

- Maps
- Tables
- Piecharts,
- Bar Charts
- Scatter Plots
- Box Plots
- GIS
- Etc....



WMA	Assessment Unit Number	Assessment Unit Name	Parameter	Station Number	Cycle 1st Listed	Designated Use	Sublist 5 Subpart (A, R, L)	Priority Ranking for TMDL
15	02040302020030-01	Absecon Creek (AC Reservoirs) (age to SB)	Mercury in Fish Tissue	Atlantic City Reservoir (Upper)	2014	Fish Consumption		Low
15	02040302020030-01	Absecon Creek (AC Reservoirs) (age to SB)	Mercury in Water Column	01410455	2010	Water Supply		Low
15	02040302020040-01	Absecon Creek (below gage)	Mercury in Fish Tissue	Atlantic City Reservoir (Lower)	2006	Fish Consumption		Low
15	02040302020040-01	Absecon Creek (below gage)	Oxygen, Dissolved	R33, 2401	2004	Aquatic Life		Medium
15	02040302020040-01	Absecon Creek (below gage)	Total Coliform	Shellfish Network	2014	Shellfish		Medium
15	02040302020010-01	Absecon Creek NB	Mercury in Fish Tissue	Atlantic City Reservoir (Lower)	2006	Fish Consumption		Low
15	02040302020010-01	Absecon Creek NB	gH	HW048639	2006	Aquatic Life		Medium
15	02040302020020-01	Absecon Creek SB	Mercury in Water Column	01410455	2006	Water Supply		Low
14	02040301160110-01	Albertson Brook / Gun Branch	gH	0140941020, 0140940070, NALDBRELL	2006	Aquatic Life		Medium
11	02040105210010-01	Absecon Ck (above 74d 55m)	gH	01461300	2014	Aquatic Life		Medium
11	02040105210010-01	Absecon Ck (below 74d 55m)	Temperature, water	01461300	2006	Aquatic Life - Trout	R	Medium
11	02040105210020-01	Absecon Ck (below 74d 55m to 118A06)	Arsenic	01461840	2012	Water Supply		Low
11	02040105210020-01	Absecon Ck (below 74d 55m to 118A06)	Escherichia coli	01461840	2012	Recreation	R	Medium
11	02040105210020-01	Absecon Ck (below 74d 55m to 118A06)	gH	01461300	2012	Aquatic Life, Aquatic Life - Trout		Medium
11	02040105210020-01	Absecon Ck (below 74d 55m to 118A06)	Temperature, water	01461300	2006	Aquatic Life - Trout	R	Medium
17	02040206060020-01	Alloway Ck (above Alloway-Woodstown Rd)	Arsenic	01482880	2006	Water Supply	A	Low
17	02040206060020-01	Alloway Ck (above Alloway-Woodstown Rd)	Total Suspended Solids (TSS)	01482880	2006	Aquatic Life		Medium
17	02040206060090-01	Alloway Ck (below HancockBrd) to Salem S	PCB in Fish Tissue	Delaware River Tribs to head of Tide	2006	Fish Consumption	L	Low
17	02040206060090-01	Alloway Ck (below HancockBrd) to Salem S	Total Coliform	Shellfish Network	2014	Shellfish		Medium
17	02040206060090-01	Alloway Ck (HancockBrd) to Newbridge	PCB in Fish Tissue	Delaware River Tribs to head of Tide	2006	Fish Consumption	L	Low
17	02040206060090-01	Alloway Ck (HancockBrd) to Newbridge	Total Coliform	Shellfish Network	2014	Shellfish		Medium



# R Automation Benefits

- Improved Efficiency
  - ~ 70% time savings for NJ
- More Time for Comprehensive Assessments
- Improved Accuracy
- Improved Validation
- More Information for Technical Decisions
- User Friendly Visualization Tools
- Overall Improved Transparency of Process

## Challenges with automation:

- Programming Skills & Time
- Validation & QA : requires scientific staff verification





# Lessons Learned

- Use the resources on the internet
- New features / packages / tools make a difference
- Train more than one person, get a computer programmer (non-environmentalist) or contract out
- It takes much longer time to program if doing it in-house
- Not a one-time deal: continuous updates needed
- IT Department approvals can be a long process
- Lookup tables – accuracy is critical
- Inconsistent data standards – continued improvement required



# Questions



## Point of Contacts:

Jack Pflaumer: (609) 633-0499

[jack.pflaumer@dep.nj.gov](mailto:jack.pflaumer@dep.nj.gov)

Roop Guha: (609) 292-1592

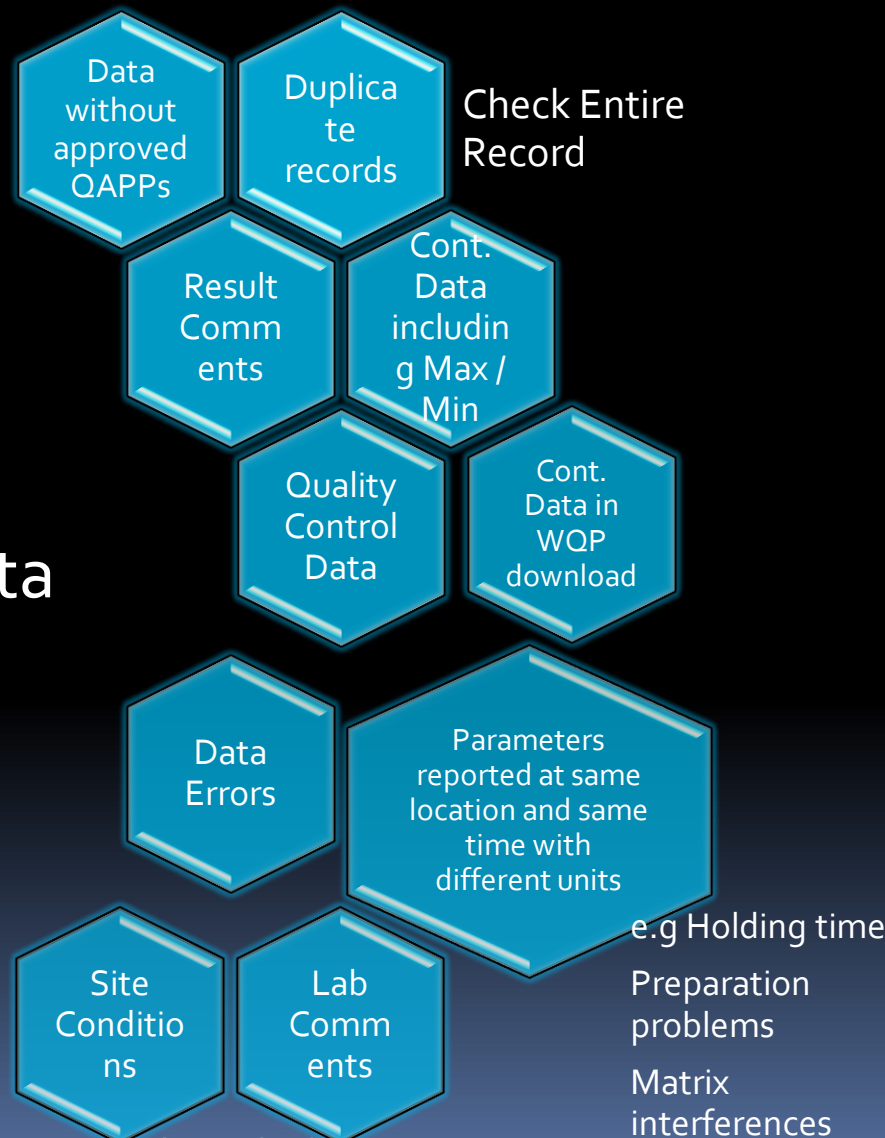
[biswarup.guha@dep.nj.gov](mailto:biswarup.guha@dep.nj.gov)

New Jersey Department of Environmental Protection (NJDEP)  
Division of Water Monitoring and Standards (WM&S)



# Quality Assurance Tools

## QA Tools Remove Data



## QA Tools Substitute Record

