NEW JERSEY’S AUTOMATED ASSESSMENT USING R

Jack Pflaumer, Biswarup Guha, Dave Rosenman

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

May 29, 2018
Overview

- Why R?
- Key Automation Steps
- Challenges and Lessons Learned
- Demonstration
Advantages of Using R Language

Challenges with automation:
- Programming Skills & Time
- Validation & QA: requires scientific staff verification
Reporting Process - Automation

Step 1. Data Assimilation
- Continuous
- Discrete
- Biological, Shellfish, Fish Tissue

Step 2. QA
- QAPPs
- Duplicates
- Data / Laboratory / Field Comments
- Validity

Step 3. Analysis
- Data Level
- Station Level
- HUC14 Level (Biological, Chemistry)
- Designated Use

Step 4. Integrated Report
- Content and Text
- Figures
- Tables

Step 5. ATTAINS batch upload + validation of upload

Automated | Manual
Legwork before Automation

Update Look-up Tables

- Parameters
  - Units
  - Names
  - New Criteria
- Stations (Continuous and Discrete)
  - Stream Classification
  - Type of waterbody
  - Colocation / station grouping
  - Representativeness
  - pH Region
- QA Comments
  - Any new comments which have not been addressed

Review QAPPs
- Whether data collected can be used for the assessment

Other Data Sources
- Identify data sources not included in WQ Portal and incorporate

Update tables + scripts (as required)
Quality Assurance Tools

- Data without approved QAPPs
- Duplicate records
- Check Entire Record
- Result Comments
- Cont. Data including Max/Min
- Quality Control Data
- Cont. Data in WQP download
- Parameters reported at same location and same time with different units
- Data Errors
- Lab Comments
- Site Conditions
- Substitute Record
- Substitute Censored Data
- Normalize reporting unit and parameter names
- Flag preliminary and estimated Data

QA Tools
Remove Data
Analysis - Station, HUC, Designated Use Levels

- Assessment History
- Change from previous cycle
- Assessment even when < target samples
- 5 & 10 year Exceedances
- Statistics: Minimum, Maximum, % Attained
- Comprehensive Assessment
- BPJ flags
- Number of exceedances, number of samples, % high exceedances
- Region

NJDEP Water Monitoring and Standards
Reporting Tools

- Maps
- Tables
- Pie charts
- Bar Charts
- Scatter Plots
- Box Plots
- GIS
- Etc....
R Automation Benefits

- Improved Efficiency
  - ~ 70% time savings for NJ

- Improved Accuracy

- Improved Validation Process

- More Time for Comprehensive Assessments

- More Information for Technical Decisions

- User Friendly Visualization Tools

- Overall Improved Transparency of Process
Challenges and Lessons Learned

Logistics
- Time
- Dedicated Resources (or lack thereof) – trained programmers preferred
- IT Department Approvals
- Functional knowledge / process familiarity
- Prepare thorough algorithm / flow chart of the process
- Inconsistent data standards – continued improvement required

Technology
- Learning R
- Continuous updates / fixes / simultaneous cycles
- New features / packages / tools make a difference
- Establish initial QA/QC process – will be adaptable process
- Lookup tables – Accuracy is critical
- Document, document, document
- Use Git / Version Control (https://git-scm.com/)
Future Projects

Cloud-based Rstudio Server hosted by EPA or some entity that can be accessed by multiple States’ water quality regulatory agencies so that the projects can be shared and possibly others could contribute – e.g. HSPF has been improved significantly because of non-funded contributions.
Resources for R/Rstudio

- **Google**
- [https://stackoverflow.com/](https://stackoverflow.com/)
- **Training - www.Datacamp.com**
- [https://www.rstudio.com](https://www.rstudio.com) (online-learning, resources/cheatsheets, products/rpackages etc)
- **Book - “R for Data Science” by Hadley Wickam** ([http://r4ds.had.co.nz/](http://r4ds.had.co.nz/))
- **Listerves / github**
  - (r_users_water_resources@listserv.state.nj.us; [https://github.com/USGS-R](https://github.com/USGS-R))
Questions

Point of Contacts:
Jack Pflaumer: (609) 633-0499  jack.pflaumer@dep.nj.gov
Roop Guha: (609) 292-1592      biswarup.guha@dep.nj.gov

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

Integrated Report Open Network, Mapping, and Assessment Navigator (IRONMAN)