Oregon DEQ Continuous data management

Finding, QCing, and managing continuous data for the Integrated Report

Travis Pritchard

State of Oregon
DEQ Department of Environmental Quality

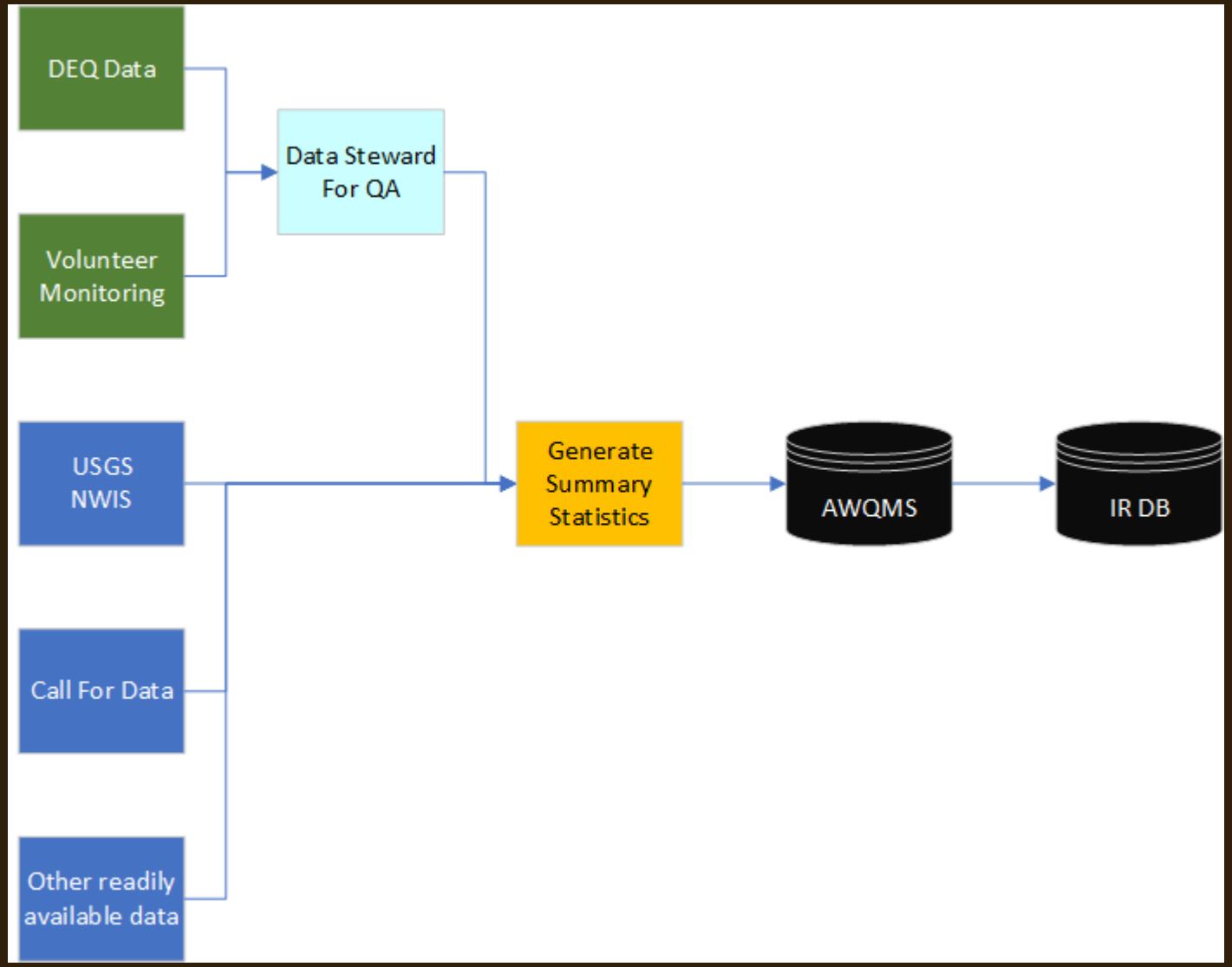
Why R?

- It’s free
- We were in a panic

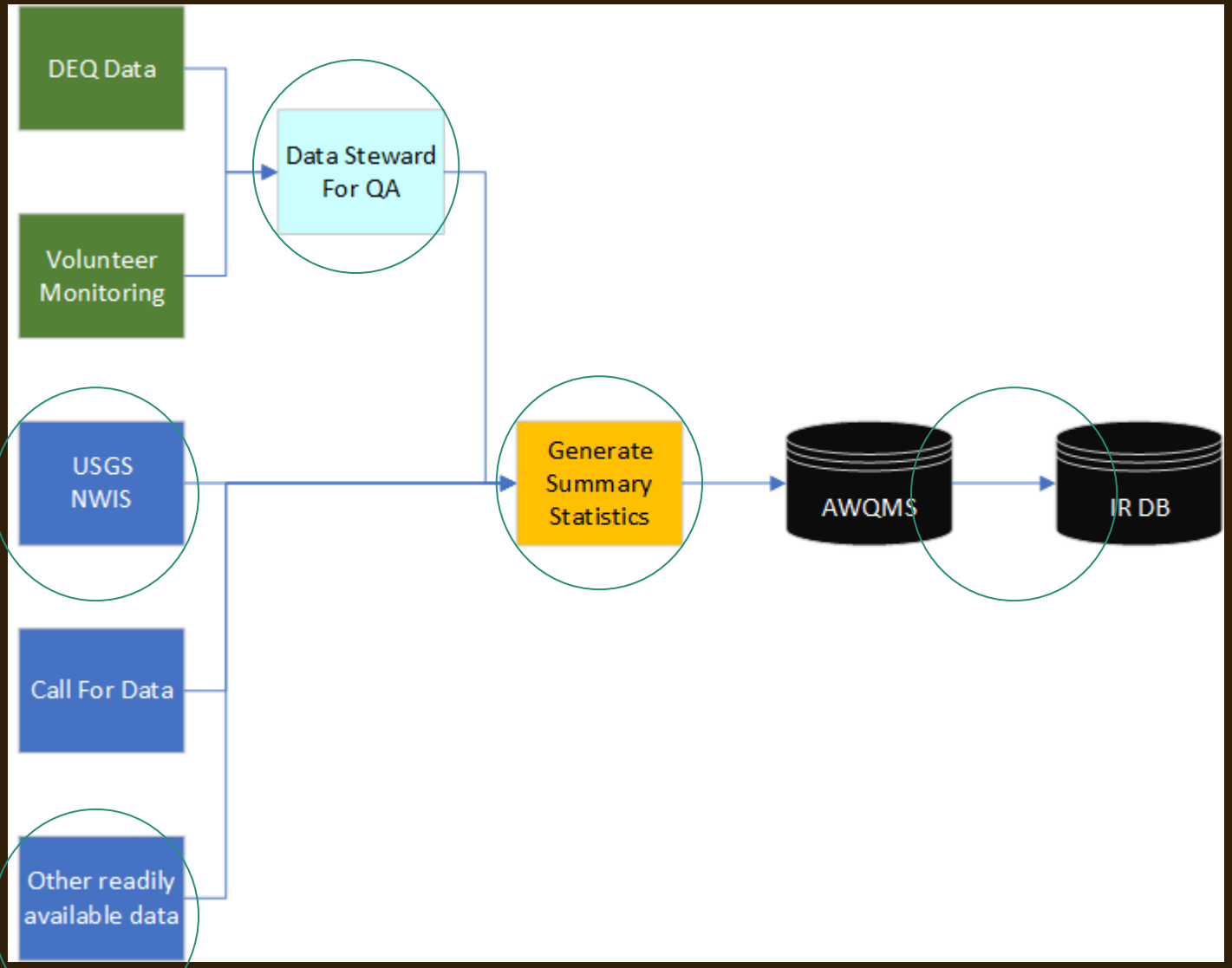
Where we use continuous data

- Temperature
  - Use summary statistics- Standard is based on 7DADM value
- Dissolved Oxygen
  - Use summary statistics- Continuous standard is based on:
    - 30 daily mean
    - 7 day mean minimum
    - Daily minimum
- pH
  - Raw continuous values
- Turbidity
  - Use summary statistics on drinking water intake data

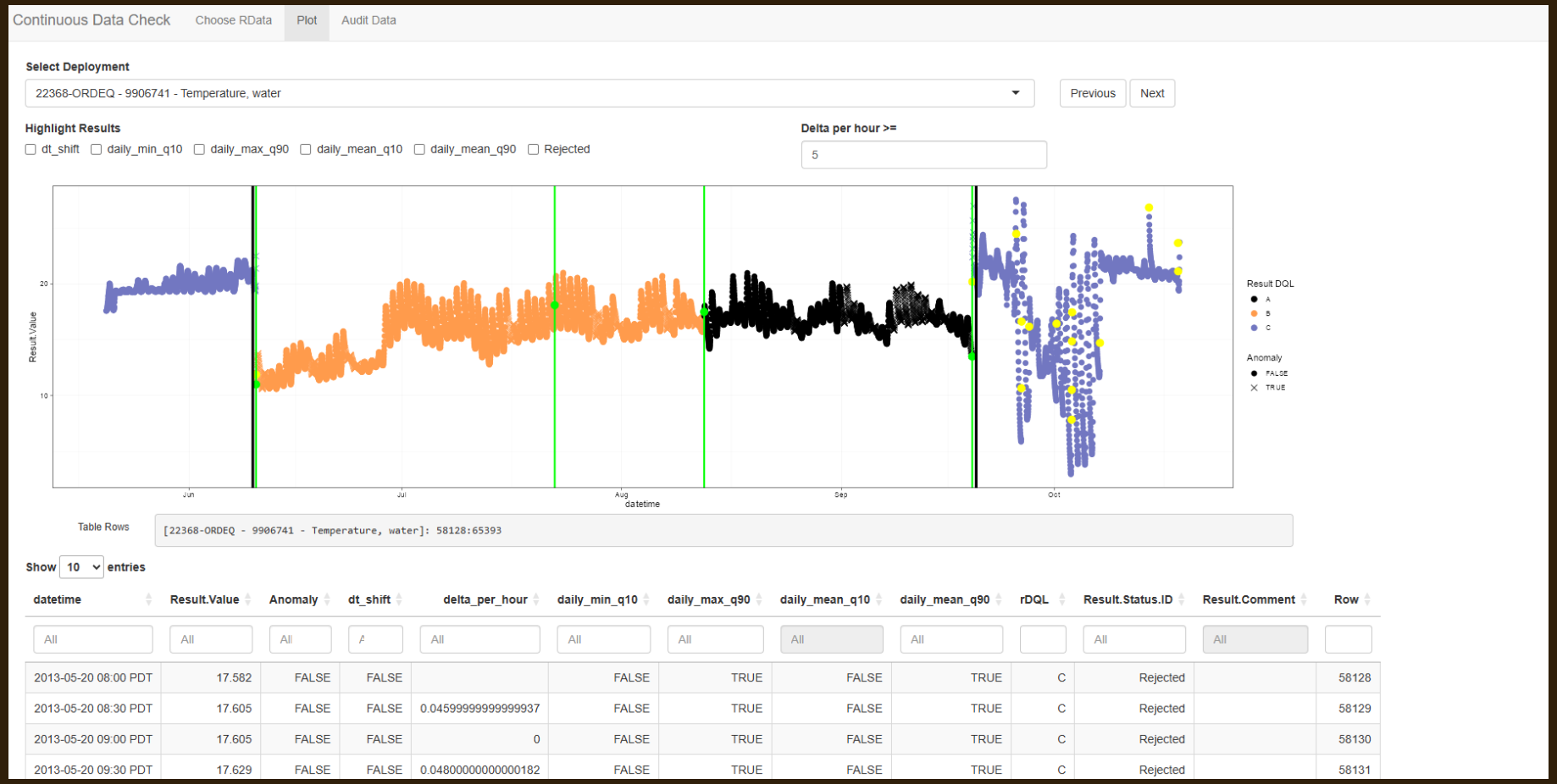
Continuous Data Management



Continuous Data Management



DEQ managed continuous QC



datetime	Result.Value	Anomaly	dt_shift	delta_per_hour	daily_min_q10	daily_max_q90	daily_mean_q10	daily_mean_q90	rDQL	Result.Status.ID	Result.Comment	Row
2013-05-20 08:00 PDT	17.582	FALSE	FALSE		FALSE	TRUE	FALSE	TRUE	C	Rejected		58128
2013-05-20 08:30 PDT	17.605	FALSE	FALSE	0.04599999999999937	FALSE	TRUE	FALSE	TRUE	C	Rejected		58129
2013-05-20 09:00 PDT	17.605	FALSE	FALSE	0	FALSE	TRUE	FALSE	TRUE	C	Rejected		58130
2013-05-20 09:30 PDT	17.629	FALSE	FALSE	0.04800000000000182	FALSE	TRUE	FALSE	TRUE	C	Rejected		58131

```r
# Setup ----------------------------------------------------------------------

#Analyst Name
analyst <- "Travis Pritchard"

#Set directory where files come from
input_dir <- ("//d[redacted]/2018/SWC_2018_Cont_Data_Sub")

#Directory where files are saved to- SHould be a voldata folder
output_dir <-"[redacted]ol_Data/Siuslaw/2018/SWC_2018_Cont_Data_Sub/R"

#Volmon template file
xlsx_input <- "WorkingCopy_Siuslaw_WC_2018_Continuous_Temp.xlsx"

#precheck file name
xlsx_pre_check_output <- "Siuslaw_WC_2018_Continuous_Temp_PRECHECK.xlsx"

# Filename for Data to load into shiny
shiny_output <- "Siuslaw_WC_2018_Continuous_Temp_SHINY_CDR.Rdata"

#Script output
xlsx_output <- "Siuslaw_WC_2018_Continuous_Temp_export.xlsx"

#changelog output
changelog <-  'Siuslaw_WC_2018_Continuous_Temp_changelog'

#- Import the Data ------------------------------------------------------------

df0 <-odeqcdr::contin_volmon_import(file=paste0(input_dir,"/",xlsx_input))

df0.projects <- df0[["Projects"]]

df0.org <- df0[["Organization_Details"]]

df0.mloc <- df0[["Monitoring_Locations"]]

df0.results <- df0[["Results"]]

df0.audits <- df0[["Audit_Data"]]

df0.deployment <- df0[["Deployment"]]
```

DEQ managed continuous QC

```
489 ####Pre CCV data missing. No DQLs get A rating-LB####
490
491 df5.results <- df4.results %>%
492
493 #'[11003-ORDEQ - 9906738 - Temperature, water]: 87191:94456 Three Rivers at Hwy 101
494   odeqcdr::dql_update(rows = c(88204:90225), "B", "") %>% #LB changed audit value
495   odeqcdr::dql_update(rows = c(91238:93050), "B", "") %>% #LB changed audit value
496
497
498 #'[22591-ORDEQ - 9906739 - Temperature, water]: 94457:98089 West Creek at HWY 101
499 #'logger error: took reading every second, only took readings on 5/20/2013
500
501
502 #'[22368-ORDEQ - 9906741 - Temperature, water]: 58128:65393 Nestucca River at Blaine Mile Post 5 downstream of Alder Creek
503   odeqcdr::dql_update(rows = c(62175:63989), "B", "") %>% #LB downgraded from A to B. Pre CCV data missing.
504
505 #'[22373-ORDEQ - 9906744 - Temperature, water]: 65394:72659 Niagara Creek at mouth
506   odeqcdr::dql_update(rows = c(66409:68429), "B", "") %>% #LB changed audit value
507   odeqcdr::dql_update(rows = c(68430:69441), "B", "") %>% #LB downgraded from A to B. Pre CCV data missing.
508   odeqcdr::dql_update(rows = c(69442:71255), "B", "") %>% #LB changed audit value
509
510
511 #'[22647-ORDEQ - 9906751 - Temperature, water]: 105356:112620 East Beaver Creek at end of Road
512   odeqcdr::dql_update(rows = c(108156:111221), "B", "") %>% #LB changed audit value. No mid-season audit- does this bring down the DQL more?
513
514
515 #'[11005-ORDEQ - 9906766 - Temperature, water]: 98090:105355 Beaver Creek at Beaver
```

Submitted data

Field 6: Checkbox (Required) with text – I certify that submitted data is of high quality and passed the applicable QA/QC protocols in the project plan or sampling plan of the submitting agency. I also grant DEQ permission to make the submitted data available to the public via DEQ's online database, AWQMS.

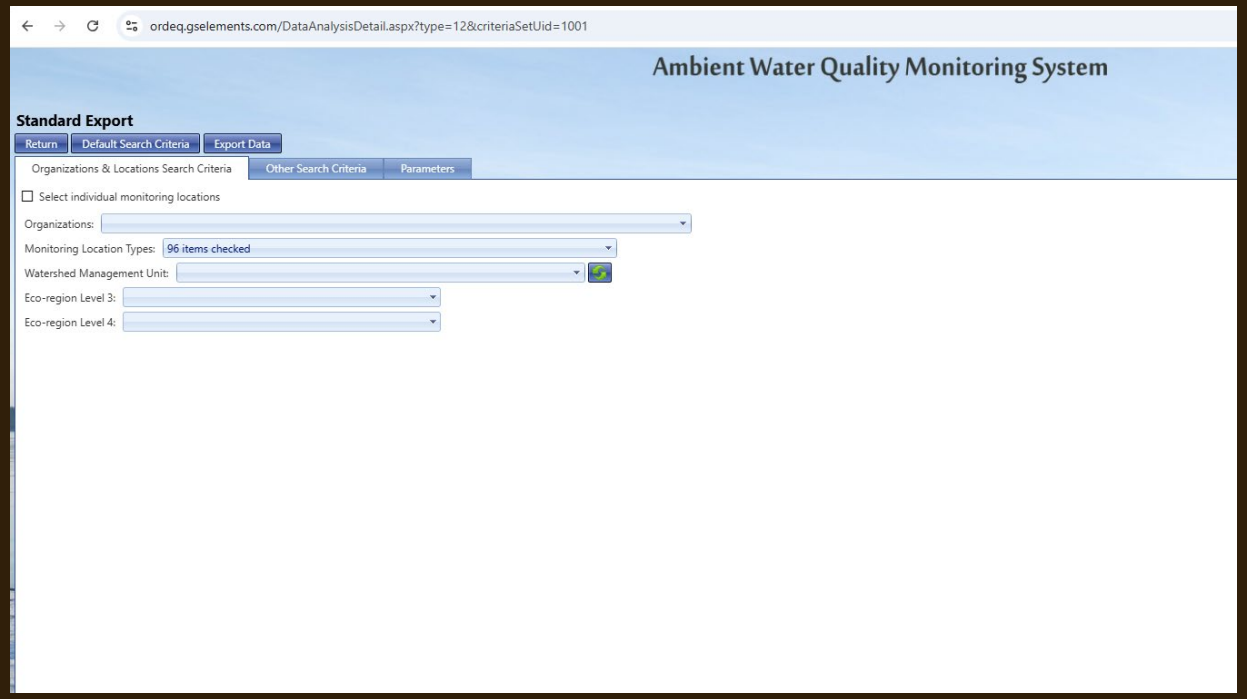
External data that we grab

- Utilize the dataRetrieval package to download continuous data from NWIS
- Only store summary statistics (except for pH)

Assessment QC

- During assessment, data is compared to the 10th and 90th percentile values for that parameter in AWQMS
- We look for and remove duplicates
- Each new category 5 listing is evaluated by assessment staff were we give a closer look at the data
- DO listings based on the daily minimum values we go back to the raw continuous data and look for reasonableness there

Getting data out of AWQMS



Usage

AWQMS_Data

Use `AWQMS_Data()` to retrieve data from Oregon DEQ AWQMS database. This dataset is too large to load into R, so you must include parameters to filter down the data. The default start date is 1949-09-15, which represents the earliest datapoint available in AWQMS.

Note - There is a shiny app to help put together this function. The app helps by allowing you to select from lists of valid values. You can clone or download the app here and run locally on your machine.

To retrieve all available data from 1/1/2017 - 12/31/2017:

```
data <- AWQMS_Data(startdate = '2017-01-01', enddate = '2017-12-31')
```

Code

- Continuous data review r package:
  - https://github.com/OR-Dept-Environmental-Quality/odeqcdr
- External continuous data pull r package:
  - https://github.com/TravisPritchardODEQ/odeqIRextdata
- AWQMS backend data package
  - https://github.com/TravisPritchardODEQ/AWQMSdata

travis.pritchard@deq.oregon.gov