

WQP Parameter and Method exploration

Harmonizing disparate data

The data from the water quality portal includes a wide range of methods and characteristic names. For example in the “chlorophyll” this can be chlorophyll a, b, or both and retrieved using a variety of methods. To know which methods and characteristic names to keep and use, we must first get a better understanding of the type of data we have.

We’ll start with the easiest first. Secchi depth

Secchi depth

Secchi Table

In many ways, the secchi disk depth measurement is the easiest water quality parameter to harmonize, because there is really only one method for measuring secchi disk depth (it’s in the name after all), and there should always be units of depth (m, ft, inches, cm, etc...). So to harmonize secchi depth measurements we simply drop all units that are not units of depth and convert all units to a single kind with a lookup table.

```
#Define a function that renames and reorders columns from the raw files
wqp.renamer <- function(df){
  simple.names <- df %>%
    dplyr::select(date=ActivityStartDate,
                  parameter=CharacteristicName,
                  units=ResultMeasure.MeasureUnitCode,
                  SiteID=MonitoringLocationIdentifier,
                  org=OrganizationFormalName,
                  org_id=OrganizationIdentifier,
                  time=ActivityStartTime.Time,
                  value=ResultMeasureValue)
  return(simple.names)
}

#Read in the raw data from '1_wqdata/tmp'
secchi <- read_feather('1_wqdata/tmp/wqp/all_raw_secchi.feather') %>%
  wqp.renamer() %>%
  #Remove trailing white space in labels
  mutate(units = trimws(units))

#Summarize by characteristic name and unit code
secchi.ct <- secchi %>%
  group_by(parameter,units) %>%
  summarize(count=n())

#Print the table
secchi.ct %>%
  knitr::kable()
```

parameter	units	count
Depth, Secchi disk depth	cm	116756

parameter	units	count
Depth, Secchi disk depth	deg C	1
Depth, Secchi disk depth	deg F	6
Depth, Secchi disk depth	ft	386537
Depth, Secchi disk depth	ft/sec	20
Depth, Secchi disk depth	in	55105
Depth, Secchi disk depth	m	1736912
Depth, Secchi disk depth	mg	187
Depth, Secchi disk depth	mi	1
Depth, Secchi disk depth	NA	12043
Depth, Secchi disk depth (choice list)		428
Depth, Secchi disk depth (choice list)	ft	1
Depth, Secchi disk depth (choice list)	m	208
Depth, Secchi disk depth (choice list)	None	13
Depth, Secchi disk depth (choice list)	NA	39281
Secchi Reading Condition (choice list)	None	643
Secchi Reading Condition (choice list)	NA	864
Water transparency, Secchi disc	in	559

Secchi disharmony

Now that we can see all the units we have we can drop non-depth units and make a lookup table to convert all units to meters.

The table below shows all the units we dropped and their counts

```
secchi.lookup <- tibble(units=c('cm','ft','in','m','mi'),
                        conversion = c(0.01,.3048,0.0254,1,1609.34))

secchi.disharmony <- secchi %>%
  anti_join(secchi.lookup,by='units') %>%
  group_by(units) %>%
  summarize(count=n())

secchi.disharmony %>%
  knitr::kable()
```

units	count
	428
deg C	1
deg F	6
ft/sec	20
mg	187
None	656
NA	52188

Secchi harmony

```
secchi.harmonized <- secchi %>%
  inner_join(secchi.lookup,by='units') %>%
```

```
mutate(harmonized_value=value*conversion)

#
#Export here.
```

Next easiest is TSS

TSS

```
#Read in the raw data from '1_wqdata/tmp'
tss <- read_feather('1_wqdata/tmp/wqp/all_raw_tss.feather')

#Summarize by characteristic name and unit code
tss %>%
  group_by(CharacteristicName,ResultMeasure.MeasureUnitCode) %>%
  summarize(count=n()) %>%
  knitr::kable()
```

CharacteristicName	ResultMeasure.MeasureUnitCode	count
Fixed suspended solids	mg/l	215914
Fixed suspended solids	mg/l	4877
Fixed suspended solids	NA	9357
Suspended sediment concentration (SSC)	%	750778
Suspended sediment concentration (SSC)	mg/l	12041
Suspended sediment concentration (SSC)	NA	3496
Suspended Sediment Concentration (SSC)	%	6758
Suspended Sediment Concentration (SSC)	g/l	7
Suspended Sediment Concentration (SSC)	mg/l	1185885
Suspended Sediment Concentration (SSC)	mg/l	1040
Suspended Sediment Concentration (SSC)	NA	5428
Total suspended solids		35
Total suspended solids	%	5072
Total suspended solids	count	1
Total suspended solids	kg	29
Total suspended solids	mg/l	2433480
Total suspended solids	mg/l	421621
Total suspended solids	None	16
Total suspended solids	NTU	1
Total suspended solids	ppm	1459
Total suspended solids	ppm	221
Total suspended solids	tons/day	32
Total suspended solids	tons/day	497
Total suspended solids	ug/l	443
Total suspended solids	ug/l	35
Total suspended solids	NA	235192

Getting harder

DOC

```
#Read in the raw data from '1_wqdata/tmp'
doc <- read_feather('1_wqdata/tmp/wqp/all_raw_doc.feather')

#Summarize by characteristic name and unit code
doc %>%
  group_by(CharacteristicName,ResultMeasure.MeasureUnitCode) %>%
  summarize(count=n()) %>%
  knitr::kable()
```

CharacteristicName	ResultMeasure.MeasureUnitCode	count
Non-purgeable Organic Carbon (NPOC)	mg/l	1393
Organic carbon	%	23844
Organic carbon	%	4890
Organic carbon	% by wt	2614
Organic carbon	% by wt	68
Organic carbon	% recovery	12
Organic carbon	count	1
Organic carbon	g/kg	7970
Organic carbon	g/kg	175
Organic carbon	mg/g	571
Organic carbon	mg/kg	2780
Organic carbon	mg/kg	656
Organic carbon	mg/l	1795644
Organic carbon	mg/l	233327
Organic carbon	None	762
Organic carbon	ppm	2855
Organic carbon	ppm	2763
Organic carbon	ug/g	67
Organic carbon	ug/kg	2
Organic carbon	ug/l	575
Organic carbon	ug/l	52
Organic carbon	NA	26879
Total carbon	%	930
Total carbon	% by wt	1457
Total carbon	g/kg	2
Total carbon	g/kg	4
Total carbon	g/m2	6
Total carbon	mg/g	14
Total carbon	mg/kg	518
Total carbon	mg/l	14692
Total carbon	mg/l	167
Total carbon	NA	28

Hardest

Chlorophyll

```
#Read in the raw data from '1_wqdata/tmp'
chl <- read_feather('1_wqdata/tmp/wqp/all_raw_chlorophyll.feather')
```

```
#Summarize by characteristic name and unit code
chl %>%
  group_by(CharacteristicName,ResultMeasure.MeasureUnitCode) %>%
  summarize(count=n()) %>%
  knitr::kable()
```

CharacteristicName	ResultMeasure.MeasureUnitCode	count
Chlorophyll	#/100ml	1
Chlorophyll	mg/l	25
Chlorophyll	mg/l	44
Chlorophyll	mg/m2	621
Chlorophyll	mg/m3	3
Chlorophyll	mg/m3	425
Chlorophyll	ml	416
Chlorophyll	None	525
Chlorophyll	RFU	2236
Chlorophyll	ug/g	1
Chlorophyll	ug/l	143415
Chlorophyll	ug/l	4364
Chlorophyll	NA	2743
Chlorophyll a		71
Chlorophyll a	g/m2	42
Chlorophyll a	IVFU	90
Chlorophyll a	mg	33
Chlorophyll a	mg/cm2	93
Chlorophyll a	mg/cm3	66
Chlorophyll a	mg/cm3	1
Chlorophyll a	mg/l	1297
Chlorophyll a	mg/l	6354
Chlorophyll a	mg/m2	15493
Chlorophyll a	mg/m2	106
Chlorophyll a	mg/m3	45308
Chlorophyll a	mg/m3	196558
Chlorophyll a	ng/cm2	1
Chlorophyll a	None	5782
Chlorophyll a	ppb	5532
Chlorophyll a	ppb	218
Chlorophyll a	ppm	803
Chlorophyll a	RFU	885
Chlorophyll a	ug/cm2	109
Chlorophyll a	ug/l	706859
Chlorophyll a	ug/l	388160
Chlorophyll a	umol/m2/s	22
Chlorophyll a	umol/m2/s	196
Chlorophyll a	NA	35067
Chlorophyll a - Periphyton (attached)	mg/m2	2545
Chlorophyll a - Periphyton (attached)	mg/m3	85
Chlorophyll a - Periphyton (attached)	ug/cm2	209
Chlorophyll a - Periphyton (attached)	ug/l	24
Chlorophyll a - Periphyton (attached)	NA	21
Chlorophyll a - Phytoplankton (suspended)	mg/m3	451
Chlorophyll a - Phytoplankton (suspended)	ug/l	2242
Chlorophyll a - Phytoplankton (suspended)	NA	314

CharacteristicName	ResultMeasure.MeasureUnitCode	count
Chlorophyll a (probe relative fluorescence)	%	49161
Chlorophyll a (probe relative fluorescence)	mg/l	30
Chlorophyll a (probe relative fluorescence)	mg/m3	860
Chlorophyll a (probe relative fluorescence)	RFU	99
Chlorophyll a (probe relative fluorescence)	ug/l	454177
Chlorophyll a (probe relative fluorescence)	volts	169
Chlorophyll a (probe relative fluorescence)	NA	25
Chlorophyll a (probe)	mg/l	1
Chlorophyll a (probe)	ppb	9553
Chlorophyll a (probe)	ug/l	54315
Chlorophyll a (probe)	volts	13415
Chlorophyll a (probe)	NA	7
Chlorophyll a, corrected for pheophytin	mg	21
Chlorophyll a, corrected for pheophytin	mg/l	2670
Chlorophyll a, corrected for pheophytin	mg/m2	11856
Chlorophyll a, corrected for pheophytin	mg/m3	71425
Chlorophyll a, corrected for pheophytin	NTU	7
Chlorophyll a, corrected for pheophytin	ppb	7678
Chlorophyll a, corrected for pheophytin	ppm	2
Chlorophyll a, corrected for pheophytin	ug/l	366621
Chlorophyll a, corrected for pheophytin	ug/ml	46
Chlorophyll a, corrected for pheophytin	NA	32455
Chlorophyll a, free of pheophytin	mg/l	23
Chlorophyll a, free of pheophytin	mg/l	22
Chlorophyll a, free of pheophytin	mg/m2	19
Chlorophyll a, free of pheophytin	mg/m2	19
Chlorophyll a, free of pheophytin	mg/m3	9436
Chlorophyll a, free of pheophytin	mg/m3	194
Chlorophyll a, free of pheophytin	mg/ml	21
Chlorophyll a, free of pheophytin	ppb	4230
Chlorophyll a, free of pheophytin	ppb	226
Chlorophyll a, free of pheophytin	ug/l	15198
Chlorophyll a, free of pheophytin	ug/l	13933
Chlorophyll a, free of pheophytin	NA	3216
Chlorophyll a, uncorrected for pheophytin	mg/l	434
Chlorophyll a, uncorrected for pheophytin	mg/m2	665
Chlorophyll a, uncorrected for pheophytin	mg/m3	19470
Chlorophyll a, uncorrected for pheophytin	mg/ml	68
Chlorophyll a, uncorrected for pheophytin	ppb	5915
Chlorophyll a, uncorrected for pheophytin	ug/l	229398
Chlorophyll a, uncorrected for pheophytin	NA	6542
Chlorophyll b		303
Chlorophyll b	mg/l	26
Chlorophyll b	mg/l	118
Chlorophyll b	mg/m2	7000
Chlorophyll b	mg/m3	18109
Chlorophyll b	mg/m3	63892
Chlorophyll b	ppb	3710
Chlorophyll b	ug/l	102211
Chlorophyll b	ug/l	15478
Chlorophyll b	NA	90240
Chlorophyll c		106

CharacteristicName	ResultMeasure.MeasureUnitCode	count
Chlorophyll c	mg/cm3	1
Chlorophyll c	mg/l	9
Chlorophyll c	mg/l	139
Chlorophyll c	mg/m3	20601
Chlorophyll c	mg/m3	69779
Chlorophyll c	ppb	3710
Chlorophyll c	ug/l	93003
Chlorophyll c	ug/l	17499
Chlorophyll c	NA	43898
Chlorophyll/Pheophytin ratio		45569
Chlorophyll/Pheophytin ratio	%	278
Chlorophyll/Pheophytin ratio	%	226
Chlorophyll/Pheophytin ratio	None	38442
Chlorophyll/Pheophytin ratio	ppb	51
Chlorophyll/Pheophytin ratio	ug/l	2862
Chlorophyll/Pheophytin ratio	ug/l	880
Chlorophyll/Pheophytin ratio	NA	94