

Check Storet data pull from 4/30/2019

Kim was asked to empty the STORET because the data was bad. It is still populated as of 5/20/2019

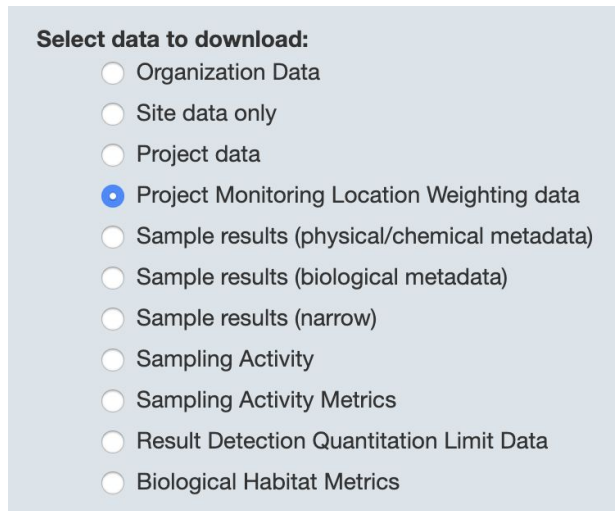
From this link I downloaded csv files for all the data:

https://www.waterqualitydata.us/portal/#organization=HUIWAIOLA_WQX&mimeType=csv

See appendix for file details

I had a note to do this with tabs vs commas next time, but I continued to use the comma separated because I didn't want to have to change any scripts, even though it might have simplified the scripts.

Here are the files I had choices of downloading.



Select data to download:

- ☐ Organization Data
- ☐ Site data only
- ☐ Project data
- ☒ Project Monitoring Location Weighting data
- ☐ Sample results (physical/chemical metadata)
- ☐ Sample results (biological metadata)
- ☐ Sample results (narrow)
- ☐ Sampling Activity
- ☐ Sampling Activity Metrics
- ☐ Result Detection Quantitation Limit Data
- ☐ Biological Habitat Metrics

Some of the choices had zero data (Project Monitoring Location Weighting data, Sampling Activity Metrics, Biological Habitat Metrics). These files were stored in /water-quality-data/storet/download-on-2019-04-30

In the data-check/results-2019-04-30-storet-check directory, I reran the script create-storet-csv-table.js against initial pull back when we discovered the problems with the storet data (after the problems with the web export file). This script makes a csv file with the data that is easier to read than storet csv files. Got the same result, So script is verified. This csv file is just used for creating an excel file for ease of viewing and documenting the problems. Created file storet-data-as-table.csv, then used it with excel to create storet-data-as-spread-sheet.xls.

Web export to storet comparison

Now checking the diffing script differences. That is the script
“find-diffs-mysql-export-vs-storet.js”

When trying to rerun it against the original physical-chemical.csv file to verify the script was running correctly, the output is a little different. There is some white space differences and I commented out the printing of this data into standard out, though it does print to standard error:

Comparison against the web export, stored in
data-check/results-2019-04-30-storet-check/diffs-web-export-to-new-storet.txt

174 total differences:

Note: there are 4 values that were qa'ed out that were then put inserted in Storet

Measurement	Number of Errors
DO	10
DO%	9
Salinity	4
Temp	30
Time	71
Turbidity	31
pH	19

There was also this difference:

These results indicate where the insitu and nutrient data has been split between two dates. One right and one wrong. These are highlighted in the storet-data-as-spread-sheet.xls

This is an example of the same data being on two different dates:

DATE DIFF SAME ENTRY WARNING OCO 20160616 20160615 NH4 2.74
 DATE DIFF SAME ENTRY WARNING OCO 20160616 20160615 NNN 5.91
 DATE DIFF SAME ENTRY WARNING OCO 20160616 20160615 Phosphate 7.56
 DATE DIFF SAME ENTRY WARNING OCO 20160616 20160615 Silicate 330.61
 DATE DIFF SAME ENTRY WARNING OCO 20160616 20160615 TotalN 78.22
 DATE DIFF SAME ENTRY WARNING OCO 20160616 20160615 TotalP 10.8
 DATE DIFF SAME ENTRY WARNING OCO 20160615 20160616 DO 6.86
 DATE DIFF SAME ENTRY WARNING OCO 20160615 20160616 DO% 101.2
 DATE DIFF SAME ENTRY WARNING OCO 20160615 20160616 Salinity 29.3
 DATE DIFF SAME ENTRY WARNING OCO 20160615 20160616 Temp 25.3
 DATE DIFF SAME ENTRY WARNING OCO 20160615 20160616 Turbidity 23.33
 DATE DIFF SAME ENTRY WARNING OCO 20160615 20160616 pH 7.96

Old storet to new storet comparison

Created a new script called find-diffs-storet-vs-storet.js to check first pull from storet in June of 2018 and pull in April of 2019. Changes all seem to be related to large numbers of significant digits that weren't accurate being rounded correctly.

Examples:

NH4	10.7358918025674	10.74 DIFF
NH4	14.6002775051814	14.6 DIFF
NH4	5.75119049718197	5.75 DIFF
NH4	6.66806803652125	6.67 DIFF

Turbidity, however went the other way on 6 entries, but still accurate.

Turbidity	0.63	0.6333333333333333 DIFF
Turbidity	2.17	2.1733333333333333 DIFF
Turbidity	2.18	2.1766666666666667 DIFF
Turbidity	2.5	2.5033333333333333 DIFF
Turbidity	3.33	3.3266666666666667 DIFF
Turbidity	5.24	5.2366666666666667 DIFF

Time changes were made on 44 entries, but as noted above, there are still 71 time errors compared to web export. I check the differences between old and new storet and all the time deltas I saw were good corrections.

Example of one that is still bad:

SampleID	OUB160616	OUB160616
Location	OUB	OUB

Date	06/16/16	06/16/16
Time	09:17	09:00 DIFF
Temp	25.2	25.2
Salinity		28.1 28.1
DO	7.00	7
DO%	103.10	103.1
pH	8.06	8.06
Turbidity		16.87 16.87
TotalN	63.47	63.47
TotalP	16.77	16.77
Phosphate		6.34 6.34
Silicate		112.01 112.01
NNN	2.64	2.64
NH4	1.79	1.79

Appendix

Data pulled in April 2019:

These files were stored in the water-quality-data repository under
storet/download-on-2019-04-30

File details

```

4813 activityall.csv
4813 biologicalresult.csv
4813 narrowresult.csv
    1 organization.csv
    2 project.csv
    31 resdetectqntlmt.csv
4813 result.csv
    88 station.csv

```

2073047590 1284751 activityall.csv

346149842 2199713 biologicalresult.csv
2506297684 1549348 narrowresult.csv
2041899127 1137 organization.csv
4094181950 473 project.csv
2221053543 5410 resdetectqntlmt.csv
3307315062 1646983 result.csv
1079383229 16706 station.csv

Result.csv is the only file that is referenced in our process. It was moved up one directory and called physical-chemical-2019-04-30.csv

These results indicate where the insitu and nutrient data has been split between two dates. One is right and one wrong.

Complete list of split insitu and nutrient data:

DATE DIFF SAME ENTRY WARNING OCO 20160616 20160615 NH4 2.74
DATE DIFF SAME ENTRY WARNING OCO 20160616 20160615 NNN 5.91
DATE DIFF SAME ENTRY WARNING OCO 20160616 20160615 Phosphate 7.56
DATE DIFF SAME ENTRY WARNING OCO 20160616 20160615 Silicate 330.61
DATE DIFF SAME ENTRY WARNING OCO 20160616 20160615 TotalN 78.22
DATE DIFF SAME ENTRY WARNING OCO 20160616 20160615 TotalP 10.8
DATE DIFF SAME ENTRY WARNING OCO 20160615 20160616 DO 6.86
DATE DIFF SAME ENTRY WARNING OCO 20160615 20160616 DO% 101.2
DATE DIFF SAME ENTRY WARNING OCO 20160615 20160616 Salinity 29.3
DATE DIFF SAME ENTRY WARNING OCO 20160615 20160616 Temp 25.3
DATE DIFF SAME ENTRY WARNING OCO 20160615 20160616 Turbidity 23.33
DATE DIFF SAME ENTRY WARNING OCO 20160615 20160616 pH 7.96
DATE DIFF SAME ENTRY WARNING OPM 20160714 20160713 NH4 2.21
DATE DIFF SAME ENTRY WARNING OPM 20160714 20160713 NNN 2.54
DATE DIFF SAME ENTRY WARNING OPM 20160714 20160713 Phosphate 4.67
DATE DIFF SAME ENTRY WARNING OPM 20160714 20160713 Silicate 136.12
DATE DIFF SAME ENTRY WARNING OPM 20160714 20160713 TotalN 79.94
DATE DIFF SAME ENTRY WARNING OPM 20160714 20160713 TotalP 13.63
DATE DIFF SAME ENTRY WARNING OPM 20160728 20160727 NH4 11.32
DATE DIFF SAME ENTRY WARNING OPM 20160728 20160727 NNN 3.19
DATE DIFF SAME ENTRY WARNING OPM 20160728 20160727 Phosphate 4.59
DATE DIFF SAME ENTRY WARNING OPM 20160728 20160727 Silicate 294.07
DATE DIFF SAME ENTRY WARNING OPM 20160728 20160727 TotalN 90.37
DATE DIFF SAME ENTRY WARNING OPM 20160728 20160727 TotalP 10.08
DATE DIFF SAME ENTRY WARNING OPM 20160922 20160921 NH4 2.35
DATE DIFF SAME ENTRY WARNING OPM 20160922 20160921 NNN 3.94
DATE DIFF SAME ENTRY WARNING OPM 20160922 20160921 Phosphate 5.33

DATE DIFF SAME ENTRY WARNING OPM 20160922 20160921 Silicate 145.77
 DATE DIFF SAME ENTRY WARNING OPM 20160922 20160921 TotalN 71.8
 DATE DIFF SAME ENTRY WARNING OPM 20160922 20160921 TotalP 14.8
 DATE DIFF SAME ENTRY WARNING RWA 20160614 20160615 NH4
 DATE DIFF SAME ENTRY WARNING RWA 20160614 20160615 NNN 21.35
 DATE DIFF SAME ENTRY WARNING RWA 20160614 20160615 Phosphate 12.38
 DATE DIFF SAME ENTRY WARNING RWA 20160614 20160615 Silicate 822.08
 DATE DIFF SAME ENTRY WARNING RWA 20160614 20160615 TotalN 84.6
 DATE DIFF SAME ENTRY WARNING RWA 20160614 20160615 TotalP 24.07
 DATE DIFF SAME ENTRY WARNING RWA 20160628 20160629 NH4 3.34
 DATE DIFF SAME ENTRY WARNING RWA 20160628 20160629 NNN 15.03
 DATE DIFF SAME ENTRY WARNING RWA 20160628 20160629 Phosphate 10.91
 DATE DIFF SAME ENTRY WARNING RWA 20160628 20160629 Silicate 633.71
 DATE DIFF SAME ENTRY WARNING RWA 20160628 20160629 TotalN 83.6
 DATE DIFF SAME ENTRY WARNING RWA 20160628 20160629 TotalP 20.43

Raw diff from web-export-to-storet differences

DO	#N/A	8.6	DIFF
DO	6.00	6.01	DIFF
DO	6.02	6.03	DIFF
DO	6.06	6.07	DIFF
DO	6.07	6.06	DIFF
DO	6.35	6.33	DIFF
DO	6.56	5.56	DIFF
DO	6.65	6.05	DIFF
DO	6.75	6.76	DIFF
DO	6.77	6.76	DIFF
DO	7.17	7.1	DIFF
DO%	100.80	100.9	DIFF
DO%	84.40	84.7	DIFF
DO%	87.10	87.4	DIFF
DO%	88.40	88.5	DIFF
DO%	92.80	92.3	DIFF
DO%	94.00	94.2	DIFF
DO%	94.20	94	DIFF
DO%	96.20	92.2	DIFF
DO%	96.70	95.7	DIFF
DO%	97.00	91	DIFF
Phosphate	10.00	10	DIFF

Salinity	#N/A	37.1	DIFF
Salinity	35.1	35.06666666666667	DIFF
Salinity	36.5	36.6	DIFF
Salinity	36.6	36.5	DIFF
SampleID	OCO160616	OCO160615	DIFF
Silicate	560.00	560	DIFF
Temp	23.8	23.85	DIFF
Temp	24.1	24	DIFF
Temp	24.6	24.7	DIFF
Temp	24.8	24.9	DIFF
Temp	24.9	25.1	DIFF
Temp	25.1	25	DIFF
Temp	25.2	25.3	DIFF
Temp	25.3	25.2	DIFF
Temp	25.5	25.6	DIFF
Temp	25.8	25.9	DIFF
Temp	26.1	26	DIFF
Temp	26.1	26.2	DIFF
Temp	26.1	26.4	DIFF
Temp	26.2	26.3	DIFF
Temp	26.4	26.5	DIFF
Temp	26.5	26.6	DIFF
Temp	26.6	26.2	DIFF
Temp	26.9	26.8	DIFF
Temp	27.1	27	DIFF
Temp	27.3	27.4	DIFF
Temp	27.4	27.3	DIFF
Temp	27.7	27.2	DIFF
Temp	27.8	27.7	DIFF
Temp	27.9	28	DIFF
Temp	28.0	28.1	DIFF
Temp	28.1	28	DIFF
Time	07:15	07:10	DIFF
Time	07:21	07:00	DIFF
Time	07:30	07:10	DIFF
Time	07:35	07:30	DIFF
Time	07:40	07:15	DIFF
Time	07:44	07:00	DIFF
Time	07:45	07:20	DIFF
Time	07:48	07:30	DIFF
Time	07:50	07:10	DIFF
Time	08:00	07:15	DIFF
Time	08:00	07:30	DIFF

Time	08:03	07:41	DIFF
Time	08:05	07:00	DIFF
Time	08:05	07:10	DIFF
Time	08:05	07:20	DIFF
Time	08:06	07:40	DIFF
Time	08:12	07:30	DIFF
Time	08:12	08:15	DIFF
Time	08:19	08:46	DIFF
Time	08:20	07:10	DIFF
Time	08:20	07:15	DIFF
Time	08:22	07:30	DIFF
Time	08:22	07:41	DIFF
Time	08:23	07:40	DIFF
Time	08:24	07:00	DIFF
Time	08:28	08:20	DIFF
Time	08:30	07:20	DIFF
Time	08:34	07:30	DIFF
Time	08:40	07:10	DIFF
Time	08:40	07:15	DIFF
Time	08:40	08:17	DIFF
Time	08:41	08:15	DIFF
Time	08:42	07:00	DIFF
Time	08:42	07:41	DIFF
Time	08:45	07:30	DIFF
Time	08:45	08:18	DIFF
Time	08:46	08:19	DIFF
Time	08:47	07:40	DIFF
Time	08:50	07:20	DIFF
Time	08:51	08:24	DIFF
Time	08:53	07:30	DIFF
Time	08:59	07:41	DIFF
Time	08:59	08:17	DIFF
Time	09:02	07:30	DIFF
Time	09:02	08:15	DIFF
Time	09:04	07:40	DIFF
Time	09:05	07:15	DIFF
Time	09:06	07:00	DIFF
Time	09:06	08:18	DIFF
Time	09:10	07:20	DIFF
Time	09:10	07:30	DIFF
Time	09:12	08:24	DIFF
Time	09:17	09:00	DIFF
Time	09:20	07:30	DIFF

Time	09:20	07:40	DIFF
Time	09:25	08:15	DIFF
Time	09:26	07:30	DIFF
Time	09:30	08:18	DIFF
Time	09:31	08:17	DIFF
Time	09:40	08:24	DIFF
Time	09:40	09:17	DIFF
Time	09:48	08:15	DIFF
Time	09:53	08:17	DIFF
Time	09:58	10:24	DIFF
Time	10:00	08:18	DIFF
Time	10:01	08:24	DIFF
Time	10:10	08:15	DIFF
Time	10:19	08:17	DIFF
Time	10:24	09:58	DIFF
Time	10:25	08:18	DIFF
Time	10:25	08:24	DIFF
Turbidity	#N/A	1.33	DIFF
Turbidity	#N/A	2.6	DIFF
Turbidity	0.63	0.6333333333333333	DIFF
Turbidity	1.51	1.44	DIFF
Turbidity	11.33	11.333333333333333	DIFF
Turbidity	2.17	2.1733333333333333	DIFF
Turbidity	2.18	2.1766666666666667	DIFF
Turbidity	2.33	2.31	DIFF
Turbidity	2.33	2.32	DIFF
Turbidity	2.35	2.38	DIFF
Turbidity	2.42	2.41	DIFF
Turbidity	2.50	2.5033333333333333	DIFF
Turbidity	2.89	2.86	DIFF
Turbidity	22.80	22.77	DIFF
Turbidity	23.30	23.33	DIFF
Turbidity	3.03	2.99	DIFF
Turbidity	3.33	3.3266666666666667	DIFF
Turbidity	3.43	3.44	DIFF
Turbidity	3.70	5.37	DIFF
Turbidity	4.33	6.77	DIFF
Turbidity	5.01	5.15	DIFF
Turbidity	5.24	5.2366666666666667	DIFF
Turbidity	5.60	5.6033333333333333	DIFF
Turbidity	5.81	5.78	DIFF
Turbidity	5.85	5.97	DIFF
Turbidity	6.34	6.3433333333333333	DIFF

Turbidity	6.36	6.37	DIFF
Turbidity	6.77	4.33	DIFF
Turbidity	7.19	6.16	DIFF
Turbidity	7.56	6.53	DIFF
Turbidity	7.76	7.65	DIFF
Turbidity	9.15	9.08	DIFF
Turbidity	9.50	9.48	DIFF
pH	7.87	7.85	DIFF
pH	7.90	7.8725	DIFF
pH	7.90	7.89	DIFF
pH	7.96	7.94666666666667	DIFF
pH	7.96	8.01	DIFF
pH	7.98	7.97	DIFF
pH	8.00	7.96	DIFF
pH	8.01	7.96	DIFF
pH	8.01	7.99	DIFF
pH	8.03	8.08	DIFF
pH	8.04	8.03666666666667	DIFF
pH	8.05	8.07	DIFF
pH	8.06	8.05666666666667	DIFF
pH	8.10	8.09	DIFF
pH	8.11	8.06666666666667	DIFF
pH	8.11	8.1	DIFF
pH	8.14	8.13	DIFF
pH	8.15	8.14	DIFF
pH	8.16	8.15	DIFF

Raw diffs from storet-to-storet differences

NH4	10.7358918025674	10.74	DIFF
NH4	14.6002775051814	14.6	DIFF
NH4	5.75119049718197	5.75	DIFF
NH4	6.66806803652125	6.67	DIFF
NH4	6.84525602725246	6.85	DIFF
NH4	6.91181871160122	6.91	DIFF
NH4	7.49025781361792	7.49	DIFF
NH4	7.49494532660023	7.49	DIFF
NH4	7.9599466144451	7.96	DIFF
NH4	8.64713601765132	8.65	DIFF
NNN	12.0877649694861	12.09	DIFF
NNN	19.0808526194758	19.08	DIFF

NNN	2.19101801350276	2.19	DIFF
NNN	3.85699401515375	3.86	DIFF
NNN	37.509237300628	37.51	DIFF
NNN	5.51707066835697	5.52	DIFF
NNN	6.93527403520151	6.94	DIFF
NNN	9.23248032076418	9.23	DIFF
Phosphate	5.4484055352494	5.45	DIFF
Phosphate	5.5089469526279	5.51	DIFF
Phosphate	9.49308746223873	9.49	DIFF
Silicate	154.016957474844	154.02	DIFF
Silicate	321.288191142931	321.29	DIFF
Silicate	591.66799095818	591.67	DIFF
Silicate	65.9028786381786	65.9	DIFF
Silicate	669.94843375276	669.95	DIFF
Silicate	701.31122517375	701.31	DIFF
Silicate	768.899396426246	768.9	DIFF
Time	07:40	08:00	DIFF
Time	07:40	08:15	DIFF
Time	07:50	08:05	DIFF
Time	08:00	08:25	DIFF
Time	08:03	08:25	DIFF
Time	08:15	07:20	DIFF
Time	08:18	07:10	DIFF
Time	08:25	08:45	DIFF
Time	08:25	08:50	DIFF
Time	08:33	08:45	DIFF
Time	08:43	09:02	DIFF
Time	08:45	09:04	DIFF
Time	08:45	09:13	DIFF
Time	08:56	09:19	DIFF
Time	08:59	09:24	DIFF
Time	09:00	09:20	DIFF
Time	09:02	09:21	DIFF
Time	09:03	09:30	DIFF
Time	09:04	09:25	DIFF
Time	09:10	09:36	DIFF
Time	09:10	09:39	DIFF
Time	09:13	09:38	DIFF
Time	09:15	09:40	DIFF
Time	09:20	07:41	DIFF
Time	09:21	08:21	DIFF
Time	09:24	09:54	DIFF
Time	09:25	08:30	DIFF

Time	09:38	10:00	DIFF
Time	09:39	09:09	DIFF
Time	09:39	09:10	DIFF
Time	09:40	08:00	DIFF
Time	09:52	10:15	DIFF
Time	09:54	11:00	DIFF
Time	10:00	08:33	DIFF
Time	10:00	09:38	DIFF
Time	10:00	10:23	DIFF
Time	10:01	10:37	DIFF
Time	10:07	10:30	DIFF
Time	10:11	07:12	DIFF
Time	10:15	09:52	DIFF
Time	10:30	10:07	DIFF
Time	10:37	10:01	DIFF
Time	10:39	10:04	DIFF
Time	11:00	07:30	DIFF
TotalN	109.540884307262	109.54	DIFF
TotalN	159.829825619469	159.83	DIFF
TotalN	62.0490239756492	62.05	DIFF
TotalN	63.5704487768097	63.57	DIFF
TotalN	64.8929180270493	64.89	DIFF
TotalN	69.2172754426557	69.22	DIFF
TotalN	73.6996269722288	73.7	DIFF
TotalN	77.7782157661093	77.78	DIFF
TotalN	87.2754175056616	87.28	DIFF
TotalP	10.6807239484954	10.68	DIFF
TotalP	12.2225648852994	12.22	DIFF
TotalP	13.3122313165983	13.31	DIFF
TotalP	14.490850109636	14.49	DIFF
TotalP	20.9176582452563	20.92	DIFF