

Interactive comment on “The National Eutrophication Survey: lake characteristics and historical nutrient concentrations” by Joseph Stachelek et al.

Anonymous Referee #1

Received and published: 21 July 2017

This is an important data set and it is good to see that it is becoming available in electronic format. The authors appear to have used good methods to bring in a large dataset. I suspect that the manual entry error rate would be close to the lower end of their error rate.

There are multiple copies of these reports floating around in various libraries. Would it improve things to scan multiple copies of the same data and check the copies against each other?

I am curious why the authors did not try to bring in the data that Stomp et al. (2011) digitized. They could then compare the two datasets. At a minimum they should try to

C1

combine the datasets, at least chlorophyll would be nice.

I am not functional in R so could not assess that part of the data product.

The data could use some quality checking. For example, there are points where phosphate exceeds total phosphorus. This is not possible.

It also would be good if the all data were all quality checked. If it takes about 1 second per data point, I calculate it would take 3 hours each for the team of authors working in pairs to check the whole thing. That would lead to a cleaner data set as well as making the error rate certain.

It would be nice to have retention all in one type of units (not years or days mixed)

Table 2. Could use the units

Interactive comment on Earth Syst. Sci. Data Discuss., <https://doi.org/10.5194/essd-2017-52>, 2017.

C2