

Citation (APA)

- Rawson, K., & Muñoz, T. (2019). Against Cleaning. In K. Rawson, & T. Muñoz, *Data-Intensive Humanities Research*. [URL](#)

Summary

- The article, *Against Cleaning* by Katie Rawson and Trevor Muñoz discusses the research process of data cleaning; defining the practice, its application in data analytics, and the potential failings of the process. Rawson and Muñoz argue that the process of data cleaning is reductive and potentially harmful to researchers, as it has the potential to frame certain aspects of data as needing to be “cleaned” to better fit into a dataset. The authors reference the *Curating Menus* research project and how the process of cleaning failed to account for many of the foods and dishes the researchers were observing, skewing the results. Rawson and Muñoz also discuss the difficulties of making datasets saleable through cleansing, proposing the concept of data indexing as opposed to cleansing.

Evaluation

- The authors of the article include Katie Rawson, the director of library resources at the University of Pennsylvania, and Trevor Muñoz, the Director of Technology in the Humanities at the University of Maryland. The authors also reference the work *On Non scalability: The Living World Is Not Amenable to Precision-Nested Scales* by anthropologist Anna Tsing. The article seeks to challenge the process of data cleaning, and as such, is biased against the practice. The authors both have backgrounds in the humanities; therefore, the proposed concept of indexing may have bias towards the field.

The authors heavily rely on and reference their own work when giving examples of the shortcomings of conventional data cleaning, opening the article up to further bias. The research was concise and exact, with the methods they used being explained sufficiently.

Reflection

- Although my knowledge of the practice of data cleansing was limited prior to reading the article, I feel as if the authors accurately defined the practice in easy-to-understand terminology. However, it should be noted that the article is heavily biased towards eliminating the practice, as it is an opinion piece. One of the major shortcomings of the article is that the authors failed to address possible counterarguments that promote the practice of data cleansing, including its use in quantitative data analytics. I feel that the author's proposal of data indexing as opposed to cleansing could potentially push the field of research forward and promote alternative data analysis methods. Overall, I feel that the authors hypothesis could show promising results if future research is conducted.