

Case study

Retraction due to noncompliance with the journal's data policy

SOURCE: "PLOS ONE retracts perfume study when data don't pass the sniff test."
<https://retractionwatch.com/2019/09/26/plos-one-retracts-perfume-study-when-data-dont-pass-the-sniff-test/>

In 2019, *PLOS ONE* published a paper titled "Social success of perfumes" by Vaiva Vasiliauskaite and Tim S. Evans of the Theoretical Physics Group and Centre for Complexity Science at Imperial College London. In the paper's abstract, the authors pointed out that they had studied "data on perfumes and their odour descriptors – notes – to understand how note compositions, called accords, influence successful fragrance formulas".

However, the paper was soon retracted by the editors of *PLOS ONE* due to concerns about the reproducibility of the study and noncompliance with the journal's data availability policy. The editors explained their concerns, writing that after publishing several questions were raised over the dataset used in the study. They noted that further inquiry revealed that it "was obtained from a third-party commercial entity" whose identity could not be disclosed "due to a nondisclosure agreement" and that the authors could not share the raw data or disclose information about how the data were collected and processed. Although the authors posted anonymized summary data on Figshare, the reported methods are insufficient "to enable other researchers to reproduce the study". The data the authors provided do not meet PLOS ONE's requirements as described in the journal's data availability policy. According to the editors, the authors admitted to them that "they cannot reproduce the analyses using another public dataset as no comparable dataset is currently available".

The authors of the paper wrote to *Retraction Watch* and explained that the data is owned by a third party and that in order to use it, they had "to agree to very tight restrictions". For instance, they pointed out that at the point of communication with *Retraction Watch*, they no longer had access to the original data. Therefore, they were "very well aware of the restrictions when writing the paper". As they wanted to be as open as possible, they made as much of the data available as they could and this data has been accessible in the repository listed in the paper's references. Before the publication of the paper, they explained the situation to the referees and to the journal. After publication, the journal reviewed the situation again and "at that point decided that the paper did not comply with their open data policy".

Questions for discussion:

- 1) Who owns the data? Do scientists have a duty to share the data? How is this duty justified?
- 2) Who is right in this debate? Are commercial interests and protection of intellectual property legitimate arguments not to share raw data?
- 3) Why might scientists have reservations about sharing their data?