

Description

Scientists publish peer-reviewed articles frequently–they make breakthroughs in research and share them with the world. These research projects also require data collection; however, usually, a bit of inference is required to connect experimental outcomes with the hypothesis. While articles are peer-reviewed, original data quality does not fall under the typical peer review process.

This is one of the questions that reviewers and peers have during pre-publication and post-publication phase respectively. The original data is important to the scientific process and enables peers to understand the thought process of researchers. Nevertheless, the nature of peer review of data sets remains largely undefined.



Dataset Peer Review: What Does It Entail?

Open data sets are simply the raw data used for analysis in research. Peer review of data sets is the process by which open data sets associated with the manuscript are assessed and reviewed. This is an important process as it both facilitates transparency as well as increases the likelihood of citation. For example, in the Bulletin of the American Meteorological Society, 11 of the 20 most cited paper were data papers. Such data papers are those whose main purpose is to present data. Additionally, due to increasing public interest in reproducing data, there is a need and responsibility among

Copyright: Enago Academy under the CC-BY-NC-SA 4.0 license

researchers to publish the data sets that are generated during experiments. Peer review of data sets can be burdensome to the peer review process, particularly considering that peer reviewers already spend 9.6 hours per review.

Furthermore, offering datasets publicly can carry some risk for the original researchers. For example, competitor researchers may prey on the data sets of others, particularly considering the competitiveness among scientists in a limited funding environment. Ultimately, peer review of data is not yet a well-defined process and, as such, the current process of peer review without the review of data requires trust in the publishing process.

Current Guidelines

There is quite a degree of variation in current standards of data peer review. For example, some journals require data sets for certain experiments, such as transcriptomics or gene expression array plates, but not of other types. Generally, societies or institutes will combine some type of external review process to accommodate for area expertise. Whereas, the repository curating the data set retains internal staff with technical expertise to review the usefulness of the data in the present form. Such technical and subject-area review includes assessment of:

- Data logic
- Consistency
- Formatting
- Non-proprietary (i.e., open-sourced/accessible)
- Plausibility
- High quality
- Handling & reuse
- Units of measurement
- Quality of collection method
- Presence of any anomalies

These checks are helpful in ensuring that the data is both of acceptable quality and readily usable. While there is great variability due to the absence of data set review guidelines and standards, most scientists do believe in data sharing. However, the process is quite informal. For example, most researchers will share data in response to a direct contact (e.g., via email or in-person at a conference). When data is shared, most investigators simply prefer formal citation as the mechanism by which the original data set is acknowledged. Even if data sets are available, they don't have to be peer-reviewed for scientists to find them useful. Though funding and publication policies currently dictate how much and where to publish such data sets, enforcing such standards has proven to be a challenge so far.

Articles Vs. Datasets for Scientists

So, what are the differences between scientific manuscripts and data sets? Manuscripts often include only portions of useful data, which are presented as graphs or figures. Manuscripts often share compelling research results that are data-driven; however, they do not include all portions of such data.

Copyright: Enago Academy under the CC-BY-NC-SA 4.0 license

Conversely, data papers are a form of data sets written as manuscripts, which include the methods and rationale of data collection. For those who are interested in research data, pure data set journals exist, which simply publish either data papers or raw data sets themselves.

Peer-reviewed articles are important hallmarks of science. This process, for all its flaws, has allowed for science to make rapid advances in the preceding decades and centuries. Peer review of data allows scientists to be certain of the methods of data collection and data quality. By extending the peer review process to data sets, scientists stand to benefit from a more solid foundation of research.

Category

1. Publishing Research
2. Understanding Reviews

Date Created
2017/06/14
Author
daveishan

Copyright: Enago Academy under the CC-BY-NC-SA 4.0 license