Practical ways to incorporate data into tenure and promotion

Jeffrey Blume

Professor, School of Data Science, University of Virginia

Recommendations

- **Track data usage**. Usage is a strong indicator of the community's interest in the data; if data are used in research, teaching, or for other purposes, the data creator should receive credit.
- **Open up avenues to demonstrate productivity**. Data do not replace articles, but they provide an additional avenue to demonstrate research contributions.
- **Provide documentation alongside the data**. Context is key to make data valuable, it is important that data are accompanied by readme files and adequate documentation.

Traditionally, it's been difficult to document dataset creation and effectively monitor the use of datasets to demonstrate their value. Similar to articles and their citations, researchers are interested in not only noting that a dataset exists, they want to show whether and how their data have been used. Jeffrey Blume has undertaken important work to adapt the tenure and promotion (T&P) policies at the School of Data Science, University of Virginia, with the goal to enable researchers to showcase contributions through the cleanup, documentation, and other steps that underpin data generation.

Faculty were broadly on board with recognizing data contributions as part of the evaluation process. Tracking the use of data outputs can be a powerful signal of their interest and value for the community. Measures of data downloads signal that the data are useful for other researchers, particularly if these are independent counts provided by a verifiable third-party. A few reliable usage indicators can be sufficient, the key is to help the researcher build a narrative about the value of their data.

The School of Data Science's guidance for applicants and review panels will include data as part of a range of open outputs considered in the T&P process, and researchers will be encouraged to report data contributions in their CVs. Similar to the school's approach to software, the expectation is for datasets to be available via a repository where they can be accessed, downloaded and cited by others. The School of Data Science will complete this change to their T&P process over the coming months and then review its implementation.