Best Practices for Data Repository Usage Statistics

Kristian Garza (1) and Martin Fenner (2)

(1) Datacite eV, orcid.org/0000-0003-3484-6875, (2) Datacite eV

This is a report on a code of practice for data usage; one of the metrics most required by the community to understand data utilization. Funders, users (i.e. researchers), and administrators of data repositories have a large interest in understanding how the datasets from these repositories are being reused. Work is underway to establish community guidelines and services for data citation, tracking and reporting dataset citations in the scholarly literature. The second most-asked for metric for datasets is dataset usage, and neither community guidelines nor tools and services are established. We are addressing this gap in the Make Data Count project, funded by the Alfred P. Sloan Foundation. The project started in May 2017, and the first issue we are addressing is to develop community guidelines to how data usage should be collected and reported. For this, we are closely working with Project COUNTER, which has established these community standards for publications (journal articles, book chapters, etc.). We will report our initial work, based on our experience applying the COUNTER code of practice for journal articles to 150k datasets from the DataONE network. We identified three major gaps when applying the COUNTER code of practice to datasets: granularity (datasets typically contain multiple components and can again be aggregated into collections), versioning, and reporting of legitimate non-human use. We will present our draft code of practice for data usage, and look forward to feedback, including any interest to adopt these recommendations.