Hannah Carter, PhD
Assistant Professor, Medicine
University of California, San Diego
9500 Gilman Drive, MC 0688
La Jolla, CA 92093-0688
Email: hkcarter@health.ucsd.edu
Phone: +1 (858) 822-4706

08/10/2020
Re: Manuscript submission to Database

Dear Dr. Cherry,

We are pleased to submit our manuscript titled "Increasing metadata coverage of SRA BioSample entries using deep learning based Named Entity Recognition" by Klie et al. for publication in Database.

In this manuscript, we address the state of metadata quality in the Sequence Read Archive (SRA) hosted by NCBI. We highlight several issues in the annotations of samples deposited in the SRA that have similarly been reported for other large data repositories. We propose a novel deep learning-based approach for metadata curation of SRA samples, utilizing Named Entity Recognition (NER) to improve the plurality and completeness of the metadata landscape. This methodology involves minimal manual curation and can be easily scaled up to broadly improve metadata quality in the SRA and other similar databases.

The current work is unpublished and is not under consideration elsewhere. All co-authors concur with the contents of the paper. The following individuals would serve as excellent reviewers: Iris Shen (Microsoft Research), Karin Verspoor (The University of Melbourne), Vickie Walker (National Institute of Environmental Health Sciences) and Mike Conway (University of Utah).

The corresponding author's email address is hkcarter@health.ucsd.edu.

We look forward to your editorial decision.

Sincerely,

Hannah Carter