Dear Editor, dear reviewers,

Our article “Using Machine-coded Event Data for the Micro-level Study of Political Violence”, which we submitted to “Research and Politics” uses the GDELT data. Since legal issues have recently come up regarding GDELT, we feel it necessary to comment on our use of the dataset.

With no official information available, neither from the official maintainer of the dataset (Leetaru) nor from formerly affiliated scholars, much speculation has emerged about the potential issues surrounding the dataset. In order to see how far these issues affect the validity of our findings or our use of the dataset more generally, we have in the past weeks sought contact with two scholars: Phil Schrodt, a former affiliate of the GDELT project, and Pete Nardulli, Professor at the University of Illinois (UoI), the suing party in the legal dispute. Our questions to both sides were (i) if there are reasons to assume that GDELT (or part of it) suffers from serious flaws regarding data quality (e.g. data fabrication) and (ii) if they would recommend pursuing our work about GDELT.

None of these scholars was allowed to comment on the precise issues behind the legal dispute. However, both assured us that data quality is not an issue; there are no indications of data fabrication. Thus, for our project this means that GDELT uses the state-of-the-art in machine coding methodology, and thus can serve as a valid test case to scrutinize this methodology.

Regarding the question of how to proceed with our project, Phil Schrodt encouraged us to go ahead and try to contact UoI to seek their consent in using the data for our work. Pete Nardulli indicated that he is not in a position to give us a formal permission (GDELT is not “his” dataset), but he encouraged us to proceed: “In my view as a scholar in this field, you would be doing the field a favor by conducting your comparison and I would love to see a copy of it when you complete it.”

In light of these reactions from respected scholars in the field and people deeply familiar with the GDELT project, we feel comfortable submitting our paper. We hope that the editor and the reviewers share our assessment of the situation. Once again, we would like to state that we are evaluating a data generation methodology, rather than a single dataset. Even if GDELT ceases to exist, there will be other datasets in its wake employing the same methodology, and our results clearly show certain deficiencies in this methodology that scholars should be aware of. This is why we believe our results are of wider interest and go beyond GDELT.

Sincerely,

The Authors.