1. a description of what data you will use

The data for our project is 9 datasets from Russian SuperGLUE benchmark which in turn were built on existing datasets from SuperGLUE or derived from them. In Russian version datasets have been either created from scratch or translated. Each of the datasets is divided into 3 subsets: training, validation and test.

2. where that data comes from, and what license it is under

The datasets can be downloaded from https://russiansuperglue.com/ webpage and are distributed under the MIT license. We do not intend to distribute them and redirect users to this source.

3. (roughly) what you will do with the data

In our project we have analysed the data and derived a rule-base approach to solve Russian SuperGLUE tasks. The output of our project is a set of shallow heuristics per each Russian SuperGLUE task.

4. whether you can/will release the data, and under what license

According to the terms of MIT license we can release the data, and the sets of heuristics will be available under the same MIT license.

5. how another research group will be able to reproduce your project

Another research group can easily reproduce our project following the README file in our github repository -- https://github.com/tatiana-iazykova/2020_HACK_RUSSIANSUPERGLUE.