# **ReadMe:**

# *Drug Orphanisation - Dataset 1: Issue-networks as excluded publics*

This dataset comprise of two .csv format files used within workstream 2 of the Wellcome Trust funded ‘*Orphan drugs: High prices, access to medicines and the transformation of biopharmaceutical innovation*’ project (219875/Z/19/Z). They appear in various outputs, e.g. publications and presentations.

The deposited data were gathered using the University of Amsterdam Digital Methods Institute’s ‘Twitter Capture and Analysis Toolset’ (DMI-TCAT) before being processed an extracted from Gephi. DMI-TCAT queries Twitter’s STREAM Application Programming Interface (API) using SQL and retrieves data on a pre-set text query. It then sends the returned data for storage on a MySQL database. The tool allows for output of that data in various formats. This process aligns fully with Twitter’s service user terms and conditions. The query for the deposited dataset gathered a 1% random sample of all public tweets posted between 10-Feb-2021 an 10-Mar-2021 containing the text ‘Rare Diseases’ and/or ‘Rare Disease Day’, storing it on an a local MySQL database managed by the University of Sheffield School of Sociological Studies (<http://dmi-tcat.shef.ac.uk/analysis/index.php>), accessible only via a valid VPN such as FortiClient and through a permitted active directory user profile. The dataset was output from the MySQL database raw as an .gexf format file, suitable for social network analysis (SNA). It was then opened using Gephi (0.9.2) data visualisation software and anonymised/pseudonymised in Gephi as per the ethical approval granted by the University of Sheffield School of Sociological Studies Research Ethics Committee on 02-Jun-201 (reference: 039187). The deposited dataset comprise of two anonymised/pseudonymised social network analysis .csv files extracted from Gephi, one containing node data (Issue-networks as excluded publics – Nodes.csv) and another containing edge data (Issue-networks as excluded publics – Edges.csv). Where participants explicitly provided consent, their original username has been provided. Where they have provided consent on the basis that they not be identifiable, their username has been replaced with an appropriate pseudonym. All other usernames have been anonymised with a randomly generated 16-digit key. The level of anonymity for each Twitter user is provided in column C of deposited file ‘Issue-networks as excluded publics – Nodes.csv’.

This dataset was created and deposited onto the University of Sheffield Online Research Data repository (ORDA) on 26-Aug-2021 by Dr. Matthew S. Hanchard, Research Associate at the University of Sheffield iHuman institute/School of Sociological Studies. ORDA has full permission to store this dataset and to make it open access for public re-use without restriction under a CC BY license, in line with the Wellcome Trust commitment to making all research data Open Access.

The University of Sheffield are the designated data controller for this dataset.