

#DataForUkraine: Adapting Social Science Tools for Crisis Response

Ernesto Calvo; Zung-Ru Lin; Donald Moratz; Olga Onuch; Graeme Robertson; Jeremy Springman; and Erik Wibbels

The full-scale invasion of Ukraine by the Russian Federation shocked many scholars of the region, international observers, and policymakers. In the initial weeks and months of the invasion, relief organizations and government agencies lacked data about events on the ground and struggled to mount an effective response. The shocking extent of violence, mass flows of people, and lack of data collection infrastructure meant that there was an urgent need for new means of rapidly producing and disseminating data. At the request of policymakers, a team of Ukraine scholars and computational social scientists constructed a social media-based event detection system providing publicly available event data in near real-time. In this reflection, we demonstrate how social scientists can provide data to decision-makers during crises. We discuss the challenges involved and illustrate the data's utility for decision-making and academic research using case studies and descriptive analysis. We hope these insights can provide guidance for researchers and practitioners as new crises emerge around the world.

Ernesto Calvo (ecalvo@umd.edu) is Professor of Government and Politics at the University of Maryland-College Park and the Director of the Inter-Disciplinary Lab for Computational Social Science.

Zung-Ru Lin (zungru@sas.upenn.edu) is a Data Scientist at DevLab@Penn at the University of Pennsylvania.

Donald Moratz (dmoratz@sas.upenn.edu) is a PhD Student at the University of Pennsylvania affiliated with DevLab@Penn.

Olga Onuch (olga.onuch@manchester.ac.uk) is a Senior Lecturer in Politics at the University of Manchester.

Graeme Robertson (graeme@email.unc.edu) is Professor of Political Science and Director of the Authoritarian Politics Lab at The University of North Carolina at Chapel Hill.

Jeremy Springman (jspr@sas.upenn.edu) is a Senior Research Associate at DevLab@Penn at the University of Pennsylvania.

Erik Wibbels (ewibbels@sas.upenn.edu) is Professor of Political Science and DevLab@Penn at the University of Pennsylvania.

doi:10.1017/xxxx

How can social scientists use their technical and substantive knowledge to support decision-makers during major crises? When the Russian Federation invaded Ukraine on 24 February, 2022, this was a question many social scientists asked themselves. The full-scale invasion shocked many scholars, international observers, and policymakers, and in the initial weeks and months of the invasion, relief organizations and government agencies lacked data about events on the ground. The extent of the violence, the scale of internal displacement, and a lack of real-time information made clear the need for the agile production and dissemination of data to address a critical need.

At the request of policymakers at the United States Agency for International Development (USAID), a team of Ukraine-scholars and computational social scientists rapidly constructed a social media-based, near-realtime event detection system to provide publicly available data on developments in Ukraine. Combining deep country-expertise with computational tools, this system distils massive volumes of information into *hourly* geo-coded data updated in near-realtime that can both inform decision-making and answer important research questions.

In this reflection, we describe our approach to identifying reliable sources of information on social media, classifying text and geotagging across multiple languages, and disseminating these data to the public. We then describe important characteristics of the data, using case studies to

illustrate how events on the ground are reflected in our data and analyzing heterogeneity in information coming from different languages and user-communities. We also discuss policy-relevant applications ranging from providing immediate relief for refugees to future war crimes investigations and demonstrate how our data allows scholars to understand how information around events spread in realtime. Throughout, we focus on the general nature of the challenges we encountered and the solutions we implemented in the hopes that it may inform analogous data collection efforts in other contexts and on other platforms.

Social Media Data for Event Detection

In March 2022, USAID approached DevLab@Penn asking for data on the timing and location of events in Ukraine that might guide the international community's efforts to support the Ukrainian government and its citizens. Doing so required building a sustainable pipeline that could produce and disseminate up-to-date data on an ongoing basis at a relatively low-cost. To accomplish this, we turned to a popular social media platform: Twitter. The popularity of Twitter among journalists, activists, and policymakers means that the platform is often the first place where events are reported and discussed. Furthermore, Twitter's API provides easy access to rich information on the content of and engagement with tweets, providing a window into the