

Big Data & Gender Data Gaps

Overview of Current Data2X Projects January 2015

Data2X supports a variety of pilot projects exploring how promising sources of big data – cell phone call detail records, satellite datasets, and social media – can help close global gender gaps. The work is exploratory and aims to demonstrate what big data can and cannot achieve. Researchers will analyze data and complete small “ground-truthing” studies where needed, and report on each of the projects’ successes and failures. Data2X will compile and disseminate the outcomes of this work into a summary report with its partner, UN Global Pulse, explaining what has been learned and ways forward for using big data to analyze gender questions. The work will be completed in 2016, although Data2X and its research partners will release preliminary findings as they become available.

Pilot Projects

Call Detail Records

Cell phones are potentially a major source of information about women’s lives. Unlocking this potential depends on **constructing models for automatically predicting the sex of cell phone users** while preserving anonymity. Data2X is funding the development of such models using datasets from various countries. Accurate sex prediction models will then allow our research partners to utilize already developed methodologies for **inferring socioeconomic welfare of cell phone users** by looking at calling frequency, size of social network, recharge frequency, and other variables related to calling behavior, and to **analyze this information by sex**. If a sex-disaggregation algorithm is possible, data on the geography of cell phone use can also be used to **reconstruct women’s mobility patterns**, which can illuminate information on access to markets, health facilities, and government services.

Satellite Datasets

Improvements in computing power are enabling analysis of the tremendous amount of satellite data presently generated. Data2X is supporting the Flowminder Foundation to create models that **utilize satellite data to increase the spatial resolution of existing information from standard surveys, such as the DHS, on key indicators** of women’s welfare, mortality, morbidity, malnutrition, contraceptive access, freedom of movement, and other topics. Such data is often only available at district or higher geographical levels; geospatial information from satellites can help generate highly localized information throughout a country. Flowminder’s work will cover a range of countries, including Bangladesh, Burkina Faso, Ethiopia, Haiti, Honduras, and Nepal.

Social Media

Data2X is also supporting a diversity of projects using social media data. Our partner United Nations Global Pulse will **analyze ideas and opinions expressed on Twitter from women all across the world on topics related to the Post-2015 Development Agenda**, and in the process, **develop a toolkit for automatically identifying the sex of Twitter users**. Data2X will also support Dr. Munmun de Choudhury of Georgia Tech University to construct **Twitter-based models to estimate regional prevalences of depression among adolescent girls in South Africa and India**. Data2X is also working with Mexico’s National Institute of Statistics and Geography (INEGI) in a related effort to **use positive and negative expressions on Twitter to analyze the mental health of adolescent girls**. Given that depression is one of the leading global causes of girls’ illnesses and little information is available on the topic, these projects hold great promise for filling a key gender data gap.

Finally, a key concern from the big data for development community relates to issues of digital access and representativeness of the datasets. Data2X will work with UN Women and UN Global Pulse to explore these issues as they pertain to women and girls, to ensure that analyses using big data take into account whose digital footprints are detected in the datasets, and importantly, who may be missed.