"Big Data" for Gender: Expanding Horizons and Recognizing Limitations

The rapidly evolving field of "big data"[1] brings myriad possibilities to measure and understand the lives of women and girls to inform effective policy interventions. Recognizing the crucial importance but also limitations of traditional data, our work beginning in 2013 has confirmed that new types of data such as satellite, digital exhaust, and internet activity data can be used to improve our understanding of the different experiences of women and girls in various geographical regions, uncover key spending patterns, and even start to glean information on mental health, a chronically under-researched area. For girls and women, the experiences of whom have often been obscured by the traditional data system, new data sources offer an opportunity to fill data gaps and bring their experiences into sharper focus. In particular, women and girls who belong to sub-populations that are difficult to enumerate, e.g. marginalized socio-economic or geographically isolated groups, may disproportionately benefit simply by being made visible by new data and technology. Importantly, "big data" also extends the boundaries and broadens horizons in terms of what topics it is possible to collect high-quality data on, making it possible to ask and answer new policy questions crucial for policy-makers of today and in the future.

However, the field of "big data" is young, and the limitations of using this type of data must be better understood. Specific research on which applications of Big Data are effective and which are not in answering gender-focused questions is scarce. Seeking to advance this body of knowledge, in 2017 Data2X launched a "Big Data for Gender Challenge", providing funding to 10 innovative projects around the world. These range from exploring urban mobility of women and girls in Santiago using call detail records (CDRs) and satellite data, to examining gender differences in access and usage of financial services in Uganda, to the combination of commercial data and longitudinal surveys to understand women's involvement in the gig economy in Mexico and Kenya. This initiative serves two purposes:

- provide an opportunity to learn more about which of the new data sources are suited to answering which types of research/policy questions;
- provide guidance on which new data sources and analytical approaches show promise for further exploration/ investment and what the limitations and potential pitfalls are that researchers and policy-makers must consider when using this type of data.

In this panel intervention, I will discuss Data2X's work in this field to date and outline the contribution it seeks to make to this vital conversation.

---

[1] Here we are using "big data" as an umbrella term referring to the large amounts of digital data continually generated as a by-product of everyday interactions with digital products or services. "Big data" is often characterized by its great volume, variety, lack of structure, and high rate of velocity. However, a universally agreed definition of "big data" does not yet exist.