



CAN BIG DATA BE USED FOR EVALUATION?

A UN Women feasibility study



ACKNOWLEDGEMENTS

We thank:

- Shravanti Reddy (Lead) and Alexandra Capello of the UN Women Independent Evaluation Service for guidance and support.
- The UN Global Pulse team for advice related to selection of data sources and facilitating access to Twitter data.
- Kristen Sample, substantive expert for global corporate evaluation of UN Women's Political Participation and Leadership (WPP) program for her insights about the WPP evaluation in Mexico.
- Andrea Azevedo, gender specialist and monitoring and evaluation professional for her insights about the campaigns in Mexico and Pakistan.
- Shazia Abbasi of Shazia Abbasi Consulting and Qamar Iqbal of Association for Gender Awareness & Human Empowerment (AGAHE) for their insights on the WPP campaigns in Pakistan.

Our advisory group:

- Michael Bamberger, Independent Evaluation Expert
- Rebecca First-Nichols, Deputy Director, Data2X
- Navin Haram, Programme Specialist, Strategic Planning, Programming and Effectiveness Unit, UN Women
- Robert Kirkpatrick, Director, UN Global Pulse
- Karin Mattsson and Roxana Flores, Mexico Country Office, UN Women
- Andrew Means, Head of beyond.uptake, Uptake Foundation
- Veronica Olazabal, Director of Measurement, Evaluation and Organizational Performance, The Rockefeller Foundation
- Papa Seck, Policy Specialist, Research and Data, UN Women
- Sangeeta Thapa, Deputy Representative, Pakistan Country Office, UN Women
- Peter York, Principal and Lead of BCT Partners' Data Analytics

Authors: Dr Claudia Abreu Lopes

Dr Savita Bailur

Giles Barton-Owen

Produced by the Independent Evaluation Service

IISBN: 978-1-63214-129-3

Lead Manager: Shravanti Reddy

Editor: David Marion

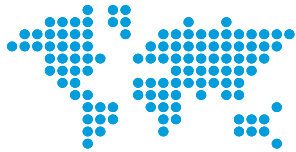
Design and layout: Yamrote Alemu H.

© 2018 UN Women. All rights reserved.

The views expressed in this publication are those of the author(s) and do not necessarily represent the views of UN Women, the United Nations or any of its affiliated organizations.

FEASIBILITY STUDY

CAN BIG DATA BE USED FOR EVALUATION?



INDEPENDENT EVALUATION SERVICE

UN WOMEN

New York, April 2018

TABLE OF CONTENTS

FOREWORD	5
ACRONYMS	6
GLOSSARY OF TERMS	7
EXECUTIVE SUMMARY	8
1. INTRODUCTION	12
1.1. Background : why big data?	13
1.2. Aims and objectives of this study	14
1.3. Pilot cases	14
1.4. Sources of big data	15
1.5. Challenges and risks of relying on big data as a source	18
1.6. Project team	19
2. METHOD	20
3. CASE STUDY DETAILS	25
3.1 Data sources	26
3.2 Pilot study in Mexico	27
3.3 Pilot study in Pakistan	27
4. DATA PROCESSING, ANALYSIS AND RESULTS	29
4.1 Twitter analysis and results for Mexico	30
4.2 Facebook analysis for Pakistan	39
4.3 Analysis of radio data	42
5. SUMMARY OF FINDINGS	43
6. RECOMMENDATIONS AND FUTURE WORK	44
APPENDIX: ALGORITHM TO DISCOVER CO-OCCURRENCES INVOLVING WORDS	46
REFERENCES	47

FOREWORD



The UN Women Independent Evaluation Service strives to continuously improve its evaluation practice and contribute to the wider evaluation community in support of gender equality and human rights responsive evaluation.

Most recently, improving our evaluation practice has meant that we need to explore how new technologies can be harnessed to improve our data collection sources and analytical capacities to deliver more credible and useful evaluations.

This study was commissioned with the above in mind to explore the feasibility of incorporating big data sources as part of the evidence base for UN Women evaluations. It represents a crucial first step toward understanding how to use big data from social media and other sources that can potentially be used for analyzing contributions to the achievement of gender equality and women's empowerment by looking at two cases – Mexico and Pakistan – related to a recent evaluation of women's political participation and leadership.

The study is an important contribution to the evaluation community's better understanding of how to make use of big data sources by providing an analysis of the pros and cons of some potential data sources, initial step-by-step protocols for their use, and recommendations based on lessons learned about using big data sources in a meaningful way for evaluation. This is an important first step on a promising method but requires further study, discussion and consideration before it can be mainstreamed as part of standard evaluation processes.

I am pleased to share this study with you and hope that you will benefit from its findings to move this area forward. I would like to take this opportunity to thank the authors and lead of the study for their hard work and contribution to improving our understanding of the possibilities for future evaluation practice.

Sincerely,

A handwritten signature in black ink, reading 'Verasak Liengsriwat'.

Verasak Liengsriwat

Director a.i., Independent Evaluation and Audit Services

ACRONYMS

AGAHE	Association for Gender Awareness & Human Empowerment
API	Application Programming Interface
DRC	Democratic Republic of Congo
FGD	Focus Group Discussions
GUID	Globally Unique Identifier
IES	Independent Evaluation Service
KII	Key Informant Interview
NLP	Natural Language Processing
ONU	Organizacion de Naciones Unidas
POS	Point-of-Sale
RCT	Randomized Control Trial
SDG	Sustainable Development Goal
SMS	Short Messaging Service
UN	United Nations
UNDG	United Nations Development Group
WPP	Women's Political Participation and Leadership
WVS	World Values Survey

GLOSSARY OF TERMS

API	Application Programming Interface (API), consisting of a set of functions, routines and protocols that define how software components communicate with each other, allowing software applications to be built from the initial code.
Construct validity	The degree to which an inference can be made from an operationalisation to the theoretical construct that it is supposed to measure (Trochim, 2006).
Interrupted time series	Research design that implies repeated measurements obtained before and after a certain period of time that is marked by one or several interventions (Ferron and Rendina-Gobioff, 2015).
Low resource languages	Less-often studied languages for which resources for natural language processing (NLP), such as machine-readable corpora, annotated data, grammar, dictionaries, treebanks and POS tagging, are scarce.
Randomized block design	Research design in which cases are sub-grouped by common attributes (forming blocks) and randomly assigned to different conditions (e.g., intervention vs control) within each block, allowing evaluating the effect of the intervention for particular blocks (Saville and Wood, 1991).
Propensity score matching	Statistical technique for observational studies that relies on a large number of covariates to predict the effects of an intervention, without the need to use a true experiment (Austin, 2011).
Nomological networks	Representation of constructs used in a study, their measurements and the relationships among them (Cronbach and Meehl, 1955).
Sentiment analysis	Technique based on natural language processing (NLP) to classify text in terms of its polarity or tonality by identifying expressions people use to evaluate or appraise persons, entities or events (Pang and Lee, 2018).
Social network analysis	Technique to represent and analyse the networks between individuals and groups (Wasserman, and Faust, 1994).
Support vector machine	Supervised learning model, with associated algorithms that analyse data used for classification tasks.
Tokenization	Process of breaking down text into its constituent elements.
Topic models	Technique based on statistical models to discover hidden topics embedded in text documents (Melville et al., 2013).

EXECUTIVE SUMMARY

The objective of this study was to investigate the feasibility of leveraging big data sources – particularly Twitter, Facebook and radio data – to improve the evaluation of gender equality and women's empowerment initiatives. In particular, this study seeks to understand the role of big data to evaluate the contribution of UN Women to women's political participation and leadership (WPP). Taking Mexico and Pakistan as two case studies, which present different challenges to access of big data sources and distinct barriers to WPP, we documented the process of accessing, analysing and triangulating big data sources with traditional data as a feasible means to provide more credible and robust insights about the effectiveness of UN Women interventions.

The community for international development evaluation has spent decades developing and refining tools for collecting and using data to determine whether social interventions work. Evaluation data often are limited to the evidence available through traditional evaluation methods constrained by rigid timeframes and scarce resources. In this respect, big data offers an additional evidence base to triangulate with and complement traditional methods. In 2017, UN Women's Independent Evaluation Service (IES) commissioned a study to determine to what extent big data could help strengthen traditional UN Women evaluations, with three key aims:

- Determine if it is possible to improve the evaluation of UN Women's work using addi-

refinement in regard to gender equality and women's empowerment issues.

- Support understanding of how UN Women and its partners might effectively use big data to support future evaluation efforts on WPP and in other thematic areas.

The feasibility study was commissioned with three main uses in mind:

1. Support UN Women IES's understanding of how and when it can incorporate big data within corporate and decentralized evaluation processes.
2. Share with like-minded organizations and partners to build learning and knowledge on the use of big data methods for evaluation.
3. Consider how to incorporate relevant data analysis and findings, specifically for evaluating WPP.

METHOD

This feasibility study focused on combining big data sources with traditional data sources to validate big data indicators. The limitations of including social

预览已结束，完整报告链接和二维码如下：

https://www.yunbaogao.cn/report/index/report?reportId=5_22321

