



# Collaborative Platform for Gender Research

---

## When **Gender** meets **Big Data**

**Marcelo Tyszler, PhD & Ewen Le Borgne**  
CGIAR Collaborative Platform on Gender Research

April 1 2019  
GRC Meeting  
Canberra, Australia

[@CGIARgender](https://twitter.com/CGIARgender)

[#GenderInAg](https://twitter.com/CGIARgender)



<https://gender.cgiar.org/themes/gender-and-big-data/>



## Overview

1. Recent activities in the collaboration
2. Key insights into the 'small grant' work
3. What's next

# 1. Recent Activities in the Collaboration

---

## Recent & on-going

### Cross-participation

- Session at the Gender Platform annual meeting in December 2017, Amsterdam
- Session at the **Big Data Platform annual meeting** in October 2018, Nairobi
- CoP SED Workshop on Data Harmonization in December 2018, Rome
  - Session on Gender and Data
  - Session on Gender at the core survey modules
- Participation in the Gender Platform annual meeting in April 2019

### Ongoing conversations

- [Slack Discussion Group](#)
- Working group within the CoP SED

### Small-grant from the CoP SED

# 1. Gender Session at the **Big Data Platform Meeting**



Collaborative  
Platform for  
Gender Research

---

## Presentation of the Gender Platform activities with Big Data

### Presentation GSMA Connected Women Program

- Mission to *reduce the gender gap in mobile internet and mobile money services in low- and middle-income*
- Gender Analysis and Identification Toolkit (GAIT): Algorithm to identify sex of the mobile user

### Exercise to gather audience interest:

*Write down one research question for scalable, replicable 'big data' research in relation to gender that you would like to see tackled in 2019*

1. How can BD help in understand how different people are differently affected or excluded from food system?
2. How can big data be used to help us understand intra-household dynamics
3. Can we use NLP (or other Big Data tool) to explore qualitative gender data in a new ways?

### *Example of interesting questions:*

- How does 'digital' work best for women in ag/livestock" / bridge gap, increase gap, etc.?
- Can big data provide better understanding of choice behavior in ag?

# F.A.I.R Data Principles

## What is FAIR DATA?



Data and supplementary materials have sufficiently rich metadata and a unique and persistent identifier.

**FINDABLE**



Metadata and data are understandable to humans and machines. Data is deposited in a trusted repository.

**ACCESSIBLE**



Metadata use a formal, accessible, shared, and broadly applicable language for knowledge representation.

**INTEROPERABLE**



Data and collections have a clear usage licenses and provide accurate information on provenance.

**REUSABLE**

## 2. Small Grant: Gender and Big Data: improving FINDABILITY of datasets

---



Starting from datasets which already are *Accessible* and *Reusable*, but have potential limitations in being *Findable* and *Interoperable*

### Initial goals:

- **Make an inventory of gender datasets** / datasets with strong gender component (quantitative or qualitative)
- **Define metadata fields** that support **gender** researchers in accessing data

## 2. Gender and Big Data FINDABILITY of datasets

---



Collaborative  
Platform for  
Gender Research

### Initial Exercise:

- **List of 61 selected Gender Datasets hosted on IFPRI Dataverse**, with direct inputs from Emily Myers (IFPRI) and Sophia Hyuer (CCAFS)
- **API-based search** (i.e. via a machine/code), **on IFPRI Dataverse using terms:** gender, women, female

## 2. Gender and Big Data FINDABILITY of datasets



### Initial Results:

	gender	women	female
Hits	57	61	27
Hits on original list	47	44	11
Misses from original list	14	17	50

- This means that when I search for “gender”, I get 57 datasets, with 47 in the original list. This means that 10 datasets were found but not provided and 14 datasets provided by IFPRI were not found.
- **How often was a dataset found:**

Found	Count
0%	6
33%	17
67%	29
100%	9

- This means that 6 datasets provided were not found by none of the searches, and 9 were found in all. The majority was found in 2 out of 3 searches



## 2. Gender and Big Data FINDABILITY of datasets



### Key words

- From the 61 datasets provided I generated a list of 217 unique keywords. From these, 8 were considered by me to be connected to gender, some very clearly, some not.

Keywords	Count
Empowerment	6
gender	37
gender analysis	2
Intrahousehold dynamics	1
WEAI	5
Women	31
Women Empowerment in Agriculture Index (WEAI)	2
women's empowerment in agriculture index	3

- Not surprisingly, “gender” is the most common keyword. There are some keywords which are not properly set and therefore not really unique, such as the 3 versions of WEAI.
- very little detailing of what, about gender, is studied in each project
- The AGROVOC keyword “sex-disaggregated data” is not used by any of the datasets

## 2. Gender and Big Data FINDABILITY of datasets

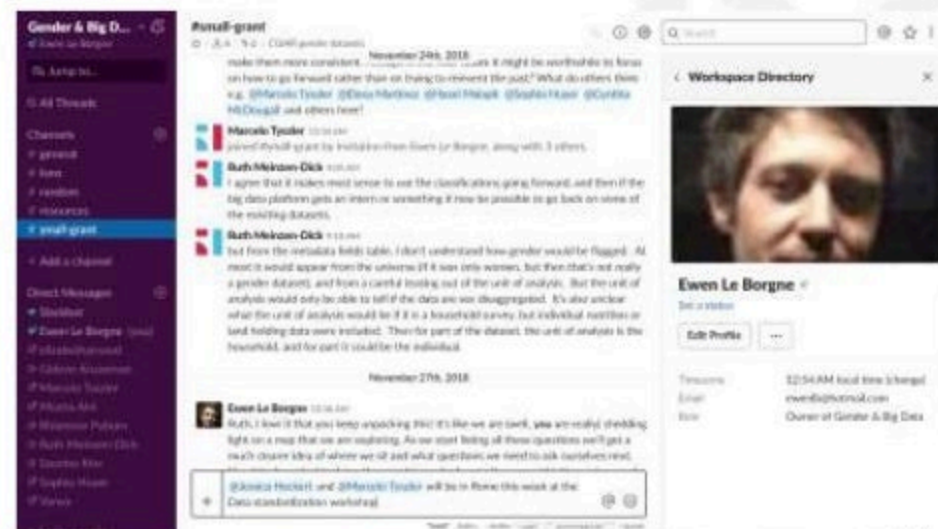


### Open questions and current answers

- **Q1. Are the datasets found but not listed indeed “gender datasets” ? Why they were not listed?**
  - This seems to be too demanding, since it involves diving into each project. An idea has been brought to “grade” a dataset in how strong the are about gender
- **Q2. How can the description of the datasets (and especially the ones not found by neither of the 3 searches) be improved so they are more findable?**
  - A simple but strong takeaway is that important keywords need to be included
- **Q3. How can information about availability of sex-disaggregated data easily (or more clearly) be incorporated in the datasets description/meta-data?**
  - The most straightforward solution now is to include “sex-disaggregated data” in the key words
- **Q4. How can we easily or more clearly incorporate information about the availability of data about different people within the same household in the datasets description/meta-data?**
  - This could either be handled in the unit of observation description or also with keywords
- **Q5. Can we be more specific in our of tagging the gender content in all datasets? How can this then be translated into keywords or meta-data fields that would facilitate searches?**
  - Similar to Q1

## 2. Slack group highlights

- Occasional updates about this work, interesting news and publications
- Conversations related to small grant and larger idea
- A space to keep everyone interested in Big Data updated
- Join the conversation!



### 3. What's next: CoP Gender and Data

- Series of Blog Posts on Gender and Big Data
- Continuation of the mini-grant:
  - Improvement of minimal advice for improvement of (gender) data repositories
  - Improvement of the list of gender research relevant keywords
  - Expansion of the gender dataset inventory by harvesting datasets from all CG centers
- Interaction on the development of the minimum WEAI, to serve as an add-on to “non-gender” research

### 3. What's next: Big Data, Big Plans

- *Where in the World are the Empowered Women?*
- Cooperation with WEAI group, CoP SED, CoP GIS
- Ideas, still under discussion:
  - Generate maps/geo-visualizations of the spatial disaggregated WEAI data (index and components)
  - Investigate if and how this correlates with spatially disaggregated data on Infrastructure and Wealth Levels
  - Explore how this can be leveraged by the use of Mobile Network Metadata (Uganda or Bangladesh)

## Looking forward, together

- Anything unclear?
- How excited are you about this (on a 1-10 scale)?
- What would be interesting research questions, activities or outputs that you would like to co-create? (10' buzzing in groups of 2-3)