



Platform for
Big Data
in Agriculture



Big Data meets Gender

Chatting up, chatting on, charging on, Changing up



40 ZETTABYTES

(40 TRILLION GIGABYTES)
of data will be created by 2020, an increase of 300 times from 2005

2020

Volume SCALE OF DATA



It's estimated that
2.5 QUINTILLION BYTES
(2.5 TRILLION GIGABYTES)
of data are created each day



Most companies in the U.S. have at least
100 TERABYTES
(100,000 GIGABYTES)
of data stored



The FOUR V's of Big Data

From traffic patterns and music downloads to web history and medical records, data is recorded, stored, and analyzed to enable the technology and services that the world relies on every day. But what exactly is big data, and how can these massive amounts of data be used?

As a leader in the sector, IBM data scientists break big data into four dimensions: **Volume, Velocity, Variety and Veracity**.

Depending on the industry and organization, big data encompasses information from multiple internal and external sources such as transactions, social media, enterprise content, sensors and mobile devices. Companies can leverage data to adapt their products and services to better meet customer needs, optimize operations and infrastructure, and find new sources of revenue.

By 2015
4.4 MILLION IT JOBS
will be created globally to support big data,
with 1.9 million in the United States



As of 2011, the global size of data in healthcare was estimated to be

150 EXABYTES
(150 BILLION GIGABYTES)



**30 BILLION
PIECES OF CONTENT**
are shared on Facebook
every month



Variety DIFFERENT FORMS OF DATA

By 2014, it's anticipated there will be

**420 MILLION
WEARABLE, WIRELESS
HEALTH MONITORS**

**4 BILLION+
HOURS OF VIDEO**
are watched on
YouTube each month



400 MILLION TWEETS
are sent per day by about 200
million monthly active users



The New York Stock Exchange captures

**1 TB OF TRADE
INFORMATION**
during each trading session



Velocity ANALYSIS OF STREAMING DATA

Modern cars have close to
100 SENSORS
that monitor items such as
fuel level and tire pressure



By 2016, it is projected there will be

**18.9 BILLION
NETWORK
CONNECTIONS**
— almost 2.5 connections



**1 IN 3 BUSINESS
LEADERS**

don't trust the information
they use to make decisions



**27% OF
RESPONDENTS**

in one survey were unsure of
how much of their data was
inaccurate

Veracity UNCERTAINTY OF DATA

Poor data quality costs the US
economy around

\$3.1 TRILLION A YEAR



The opportunities?

CGIAR Collaborative Platform for Gender Research



My Sites CGIAR Gender Customize 270 + New Edit Page

Howdy, genderplatform

The opportunities?

Big data presents various opportunities that should be approached as strategic conversations by the gender research community. Here are some emerging conversation starters. Warm thanks to Leland Glenn and colleagues at Pennsylvania State University for sharing very useful thoughts on this topic.

Big data shapes the framing of research questions among funding agencies and research institutes

Big data is the talk of (donor) town

This trend means that big data cannot and should not be ignored, or the rules of the game will be changed against an informed opinion by gender researchers. Whether a trendy 'buzz' or a longer term approach to work, big data is influencing the behavior of major players.

Big data may well favor quantitative research based on large data sets – but what is the place for and value of interpretive and contextualized research focusing on smaller data sets?

Small is beautiful!

Big data creates opportunities to analyze much larger data sets. The risk is that indeed funding agencies and research organizations tend to favor research based on larger data sets and find it increasingly difficult to justify research that looks at small data sets in a given context and tries to elaborate a more ethnographic approach towards understanding gender relations. The gender research community has much to say about this and should defend its stakes. Some other thoughts for consideration: if gender research data sets were interoperable and for instance interviews transcribed, the gender community would end up with a gigantic dataset; A lot of non-gender research which has sex-disaggregated data can also lead to a gender analysis; Organizations like Data2x, the International Food Policy Research Institute (IFPRI), The International Center for the Rights of Women (ICRW) do have some good quantitative datasets that could be mobilized and better linked to big data analyses.

Opportunities (2)

CGIAR Collaborative Platform for Gender Research



Is transparency and open data good everywhere, all the time, for everyone?

Mind your own business

Big data assumes that it is. Gender researchers know that there are real ethical risks associated with making small scale data too widely accessible, to the extent that the people concerned in any data set can be recognized and tracked down. Following a longer tradition of 'Do no harm', the gender research community has to inform where, when, and how data can and should be made widely accessible, and when not because the life of some people may depend on it. Where can ethical boundaries be laid to safeguard gender research in the age of big data?

No FAIR data standard is in place for gender research (yet)

Not FAIR!

FAIR stands for **F**indable **A**ctionable **I**nteroperable and **R**eusable. Everyone agrees on the principle of FAIR data. But practically this means paying attention to the way data is tagged, stored and shared with other repositories and users. That practical hurdle is still too big for many researchers, but there lies a real opportunity to make gender research benefit from the sirens of big data.

Isn't there a middle way (between latest data-crunching technology and no-nonsense ethnography, between large and small data sets, between open / transparent and semi-open / protected)?

The ultimate opportunity is to inform the debate around the potential of big data so that it takes into account the best methods of conducting research depending on the context and purpose at hand. Most intelligent innovations integrate existing options rather than replace them. Gender scientists can support the growth of big data with evidence of what research methods work and which don't, in their field.

**Don't like the conversation?
Change it!**

Work so far?

CGIAR Collaborative Platform for Gender Research



- Last year's session (with Gideon →)
- Initial small group (SHEA) with us
- Rhiannon joining the Big Data Steering Committee
- Some work on a 'small grant' concept note
- Soon... platform transition?



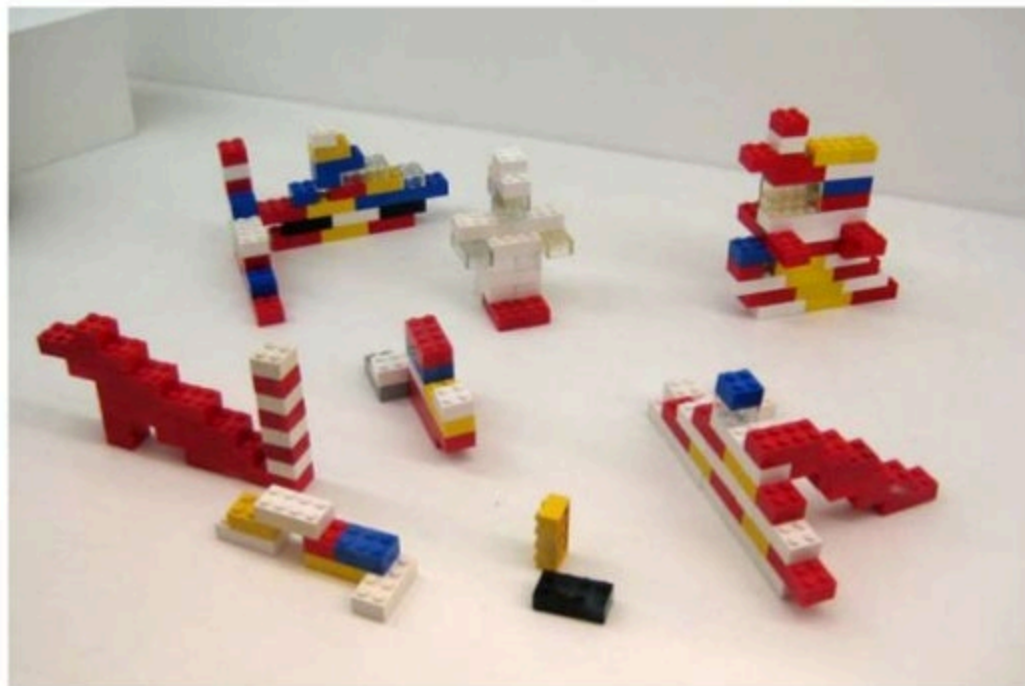
Platform for
Big Data
in Agriculture

The small grant

CGIAR Collaborative Platform for Gender Research



- Compatibility issues?
- Useful meta data?
- ...Leading to useful common research questions?



Who's helping us?

CGIAR Collaborative Platform for Gender Research



RESEARCH
PROGRAM ON
Agriculture for
Nutrition
and Health

Led by IFPRI



PennState



Platform for
Big Data
in Agriculture



Next steps

- This session!
- The Big Data Convention
- The Small Grant work
- The Inspire Challenge...

CGIAR Collaborative Platform for Gender Research



So... What?

This session

CGIAR Collaborative Platform for Gender Research



- Which data sets can we count on?
- What meta data can we work with?
- What principles should we work with to manage/share data?