# **Viva la revolución?**
## Harnessing the Data Revolution for Good

Eva.Jespersen@undp.org, @hdr2015, May 2015
Human Development Report Office

# 'Houston [HQ] we have a problem'

**POOR NUMBERS**

MORTEN JERVEN

HOW WE ARE MISLED BY AFRICAN DEVELOPMENT STATISTICS AND WHAT TO DO ABOUT IT

- About half of **children** sub-Saharan **Africa are not registered at birth**

- Low **capacity** to produce, coordinate and communicate **official statistics**;

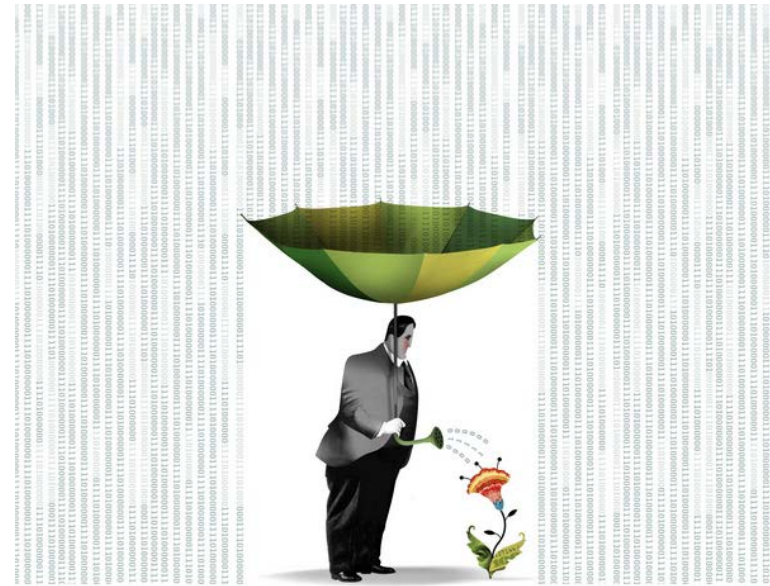- Country MDG data annually, but much is **extrapolated**

**Data Deluge or data drought ?**

... A WORLD THAT COUNTS

# A Wealth of New Data

- ICT is fueling a **new world of data**

- often **people-generated** transactional data (mobile phones, credit card, social networks)

- … and **privately-owned**

- '**big data**' / '**small data**'

**"Getting information off the Internet is like taking a drink from a fire hydrant."**

*Mitchell Kapor*

# Putting new data to good use

**Mobile data** helped **report 18 million births** in Nigeria in 2011-12

**SMS surveys** helped **reduce malaria medicine stock-outs:80%** in **Uganda**

**Google search** data may **predict** everything from recessions to flu epidemics

# More Data; More Possibilities

- **Massive samples** offer extremely fine granularity

- Data available in **real time** (*nowcasting*)
- Potential for real-time policy making

- A **window** into how **people behave** (not how they report they behave)

- **Humanitarian benefits** too e.g Facebook Safety Check in Nepal

# Considerable Strengths

# Increased Demands

- **Gaps remain** in **monitoring 8 MDG** using official statistics

- **More gaps** to come with **data-ambitious (17) SDGs (169/370)**

- Increased C**itizens demand** for information **to hold authorities accountable**

Could The **Data Revolution** solve some of these **problems** …

THE WORLD WE WANT

… A WORLD THAT COUNTS

# (IEAG) Data Revolution for Sustainable Development

- The integration of new data with traditional data for more **quality, detailed, timely and relevant** information;
- Greater **openness and transparency**, without invasion of privacy and abuse of human rights
- **Minimising inequality** in production, access to and use of data;
- More empowered people, **better policies, and decisions**, participation and accountability

# New Energy; Wider Community

- **New energy and broad excitement** at the potential
- Many statisticians are – at best – **cautiously optimistic**
- Some **antagonism** between **official statisticians** and **new data providers**, especially when NSOs receive **no credit**
- Challenges to **integrate new data into the old statistical system**

# Data Revolution Challenges

- **Data sets are massive** but can be
  - massively **biased** (selection bias etc.)
  - massively **hard to manage or analyse, and store**.

- A new "big data" science is emerging... but **little capacity in NSOs** so far..

- Big data can create big distractions: managing a world of information overload?

## Visible weaknesses

- **Expanding statistics**
  - Resource demands, cheaper techniques but demands on data management and analysis
  - Data often owned by large international operators
  - Commercial value (beyond humanitarian)
  - Legal issues
  - Ethical concerns
    - Applying fundamental principles for official statistics

## Visible weaknesses

# Data Revolution - Threats

- Concerns include **protecting privacy** & ensuring **data are used only for good**

- Already **arguments between NSOs and big data users & providers** (Tanzania)

- New collaboration (e.g. DANE Colombia)

- Potential for **worsening inequality of information**

# Global Cooperation

# Next ...

- Invest in data, capacity  building and statistical literacy
- Strengthen independent trusted statistical office that can stay relevant in a changing world
- Complementarity not antagonism between public and private providers
- Regulate for privacy and data rights
- Public data should be 'open by default'

# Big Data is still that in Africa

# The Data Revolution

IIIOIOOIIOO
OOIIOO **DATA**
**REVOLUTION**
O **GROUP** IIOOI
OIIOOOIOOII
IIIIOIIIOOOO

**1** **DIGITALLY GENERATED**
Data is created digitally, not digitized manually, and can be manipulated by computers.

**2** **PASSIVELY PRODUCED**
Data is a by-product of interactions with digital services.

**3** **AUTOMATICALLY COLLECTED**
A system is in place that automatically extracts and stores the relevant data that is generated.

**4** **GEOGRAPHICALLY OR TEMPORALLY TRACKABLE**
For instance, this is the case in mobile phone location data or call duration time.

**5** **CONTINUOUSLY ANALYZED**
Information is relevant to human well-being and development, and can be analyzed in real time.

... A WORLD
THAT COUNTS

# Inequality and Invisibility

- **Growing inequalities** in being able to access or analyse data;

- Exclusion from lack of resources (inc. capacity): "**information rich or poor**"

- The **invisible** are barely covered by existing data

# Private Sector

- Data has been **shared for humanitarian purposes but less often shared for general statistics**

- Data are **expensive to collect and commercially valuable**. Persuading **companies to share data as a public good**?

- **Ethical concerns** over data too. Do we need **fundamental principles for all statistics** (not just official statistics)?

# National Statistical Offices

- Consider the **stereotype of a conservative NSO**, not willing to react to urgent demands.  But for good reason: **reliable statistics take time**.  The **revolution could change this**.

- Yet **greater analytical capacity, and more experience of Big Data are required**: NSOs are **right to be reluctant to rely on new data**.  But they must take the revolution seriously or risk becoming irrelevant.

- NSOs to **leapfrog their statistical portfolio** and **reduce** reliance on expensive (+ difficult to run) sample /s