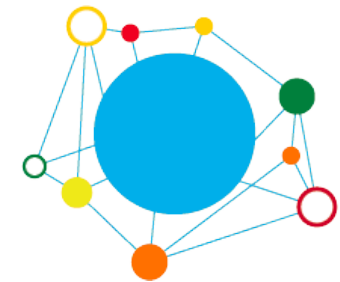


IX NIC.br Annual  
**Workshop  
on Survey  
Methodology**

Data for public statistics:  
Data Science, Big Data & Artificial Intelligence  
20-21-22-23 May 2019  
Brazilian Network Information Center – NIC.br  
São Paulo, Brazil



**DATA-POP**  
ALLIANCE



# **Make Measurement Matter: Big Data and Artificial Intelligence for Monitoring and Promoting Sustainable Human Development**

Emmanuel Letouzé, PhD

Director, Data-Pop Alliance | Director, OPAL Project

Visiting Scholar, MIT Media Lab | Connection Science Fellow, MIT



Eurostat New Techniques and Technologies  
for Statistics Conference  
Brussels, March 14, 2019



**MIT**  
**Connection**  
**Science**

Part 1:  
**Genesis, Context, Concepts  
and Questions of the 4<sup>th</sup> Industrial  
'Data' Revolution**

Part 2:  
**Statistical Measurement and  
Sustainable Development in the  
Age and Big Data and AI**

Part 3:  
**Pillars and Pathways of a People-  
Centered, Data-Enabled Human  
Development Revolution:  
Towards and Human AI**

Part 1:  
**Genesis, Context, Concepts  
and Questions of the 4<sup>th</sup> Industrial  
'Data' Revolution**

Part 2:  
**Statistical Measurement and  
Sustainable Development in the  
Age and Big Data and AI**

Part 3:  
**Pillars and Pathways of a People-  
Centered, Data-Enabled Human  
Development Revolution:  
Towards and Human AI**

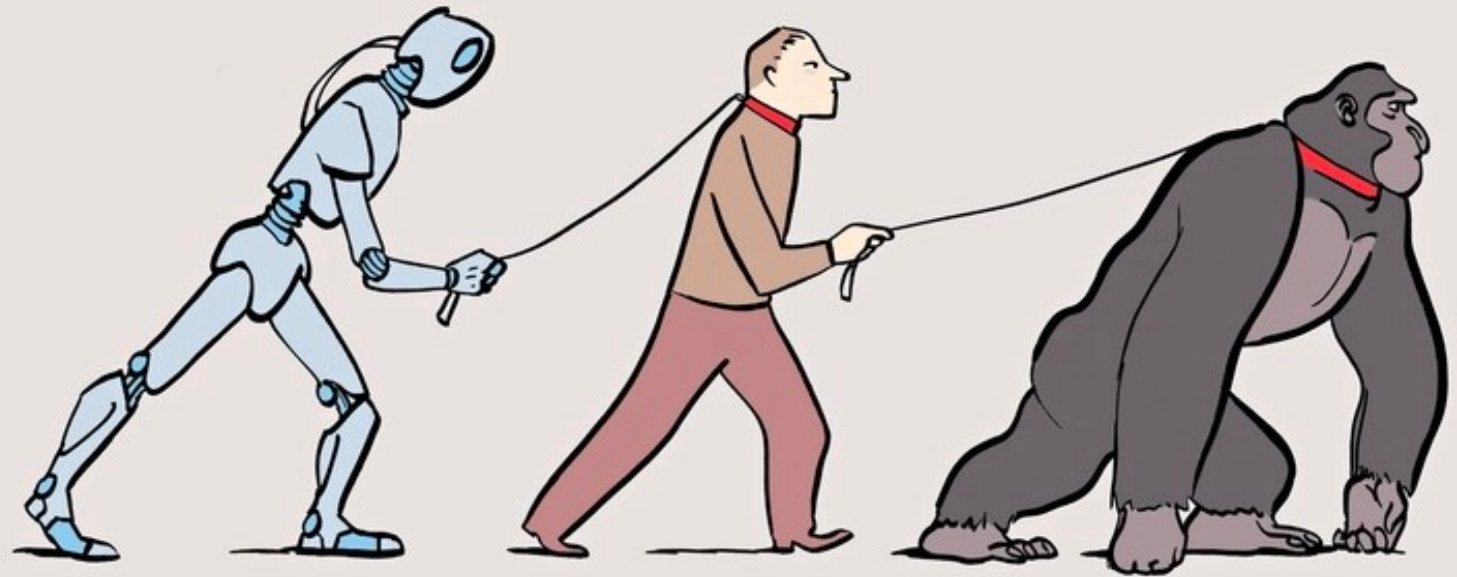
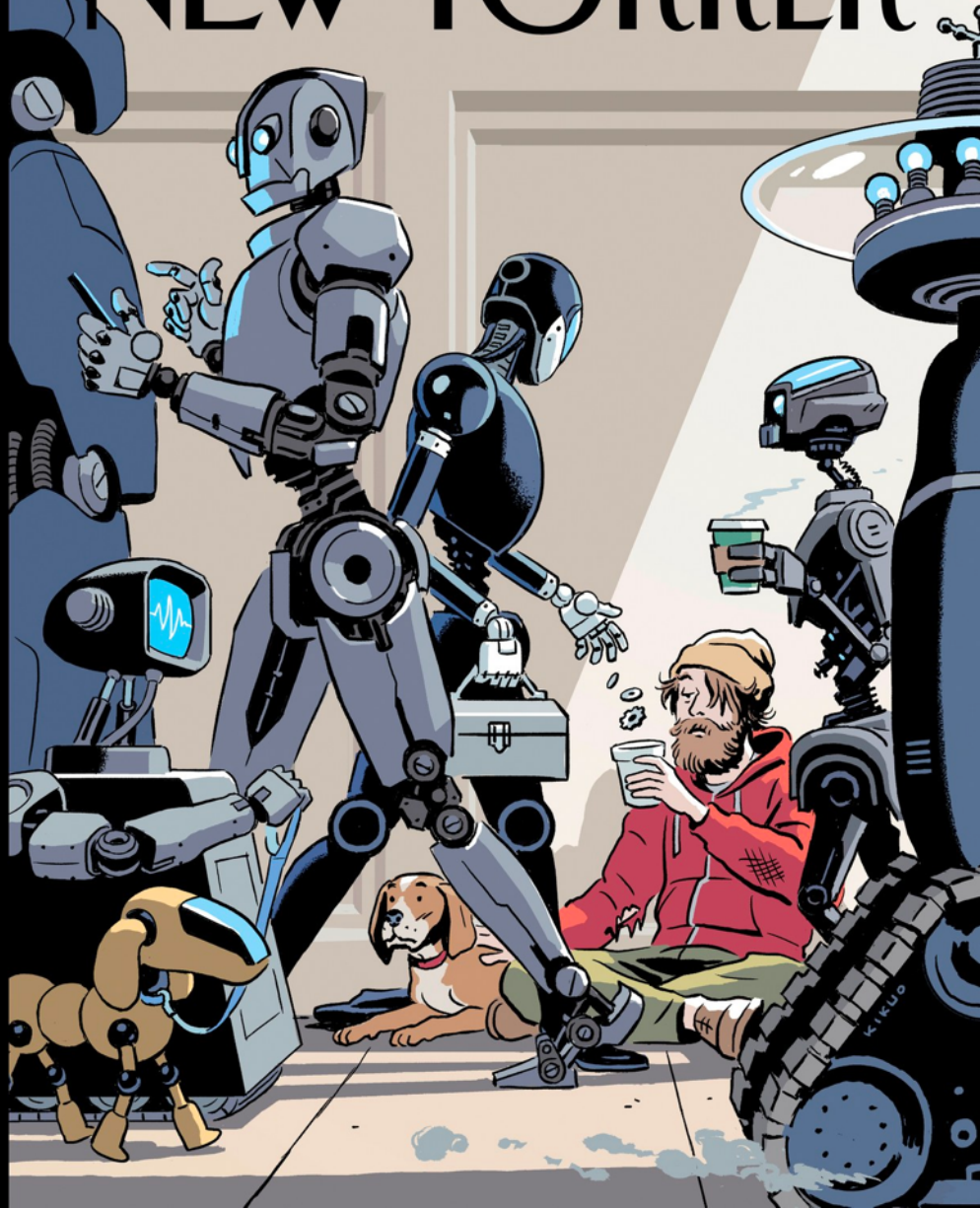


PRICE \$8.99

THE

OCT. 23, 2017

# NEW YORKER



1. Are we ever going to be enslaved by AI-powered machines? Be discriminated by algorithms? Lose our jobs? Have a machine-driven war? Or all get paid to do no work? Hopefully and probably none of the above. But...
2. Can we envision and build a better world where humans and machines cooperate – and where measurements and facts matter for sustainable human development? a "Human AI" or "human-machine ecology"? What would it feel like, look like, and take? Where are we now and can we go?

THE DATA  
REVOLUTION  
IS HERE!

The good news is we can now  
measure your poverty levels  
at amazing levels of geographic  
granularity in real time!

The bad news is we still  
can't do anything about it



MANU NITS  
2019



HAL VARIAN PREDICTED:  
"THE SEXIEST JOB OF THE 21<sup>ST</sup>  
CENTURY will BE STATISTICIAN!"

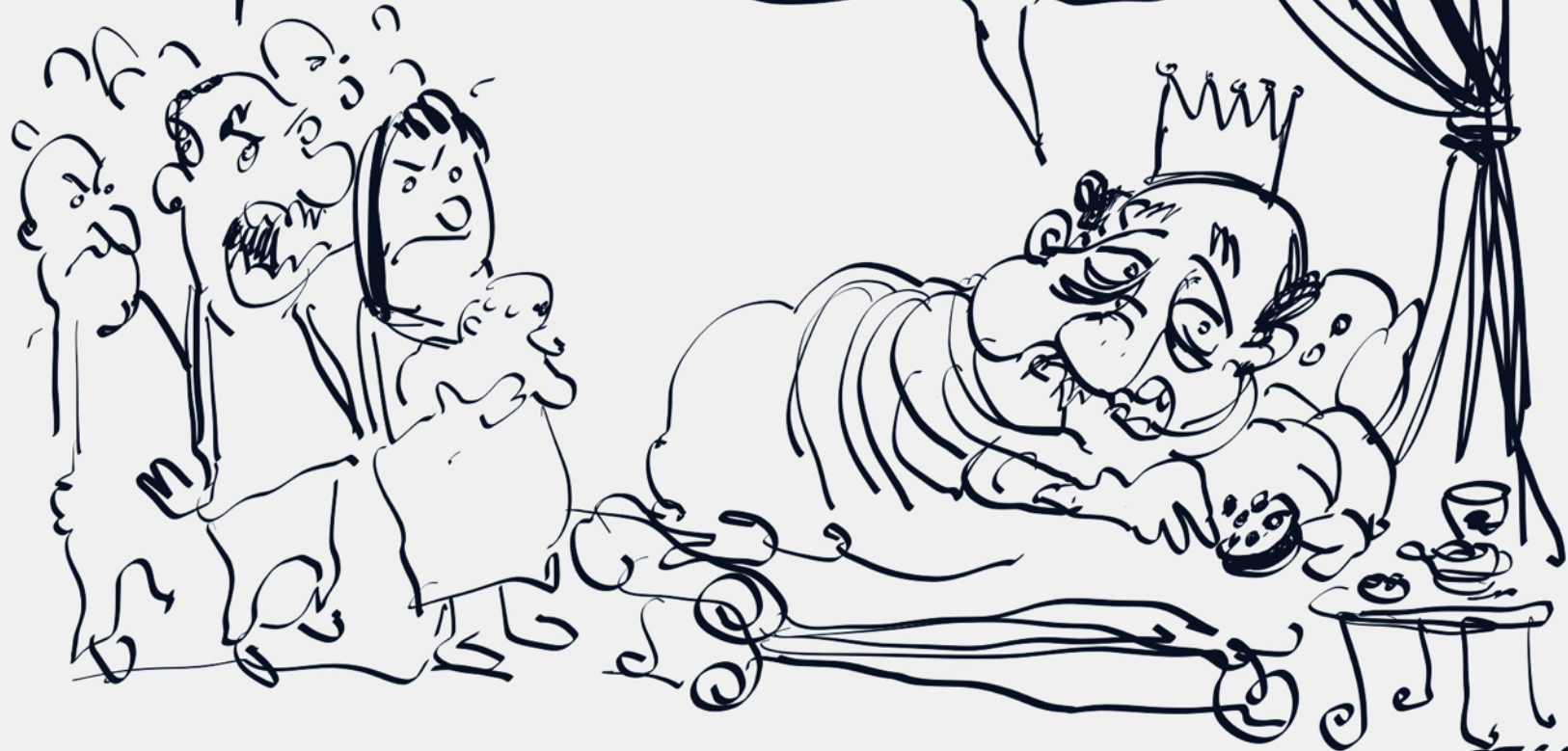
Note that he didn't specify  
when in the 21<sup>st</sup> century...



MANU  
NTTS 2019

SIRE! WE THINK YOU  
NEED BETTER DATA!

i THINK I NEED  
BETTER DATES..



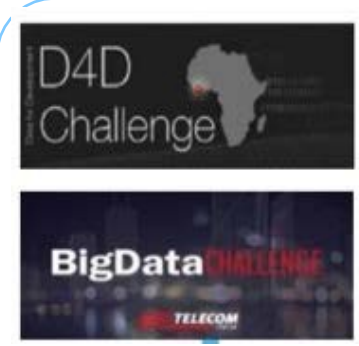
MANU NITS 2019



# A decade of “Data Revolution”; a decade until 2030: expectations, experimentations, controversies, slow changes...



*“We are at the beginning of what I call **The Industrial Revolution of Data.**”*  
 Joe Hellerstein ,  
 Nov. 2008



**Fix Africa’s Statistics**  
 By Marcelo Giugale

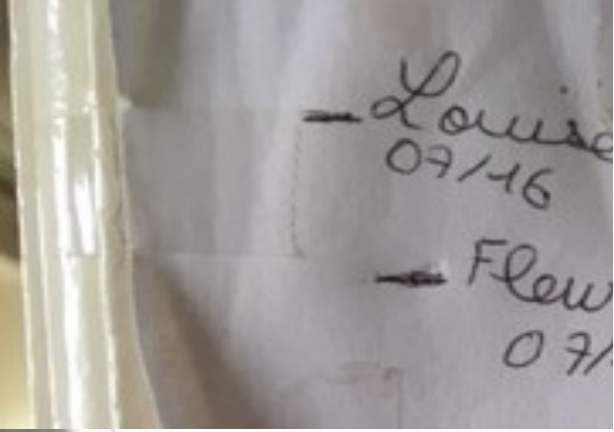


Data and development  
 Off the map **The Economist**  
*Rich countries are deluged with data; developing ones are suffering from drought*

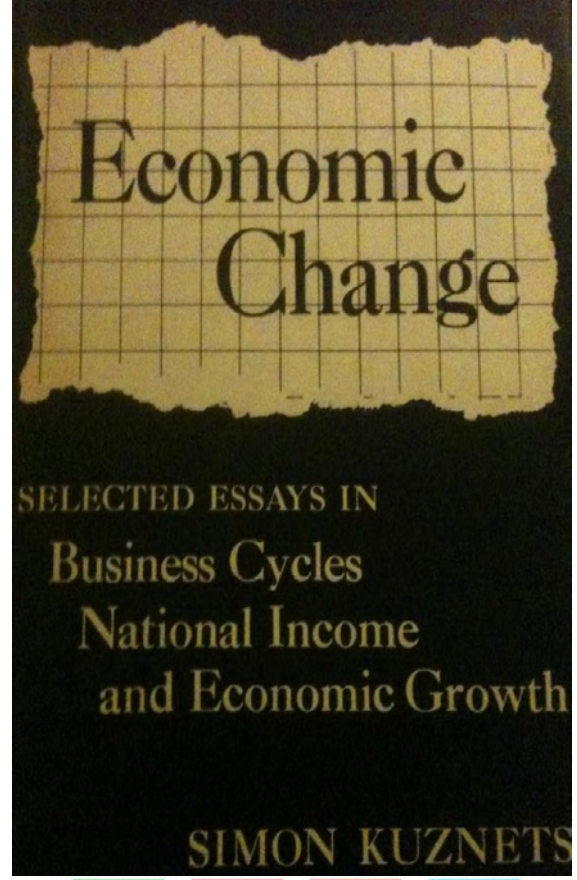




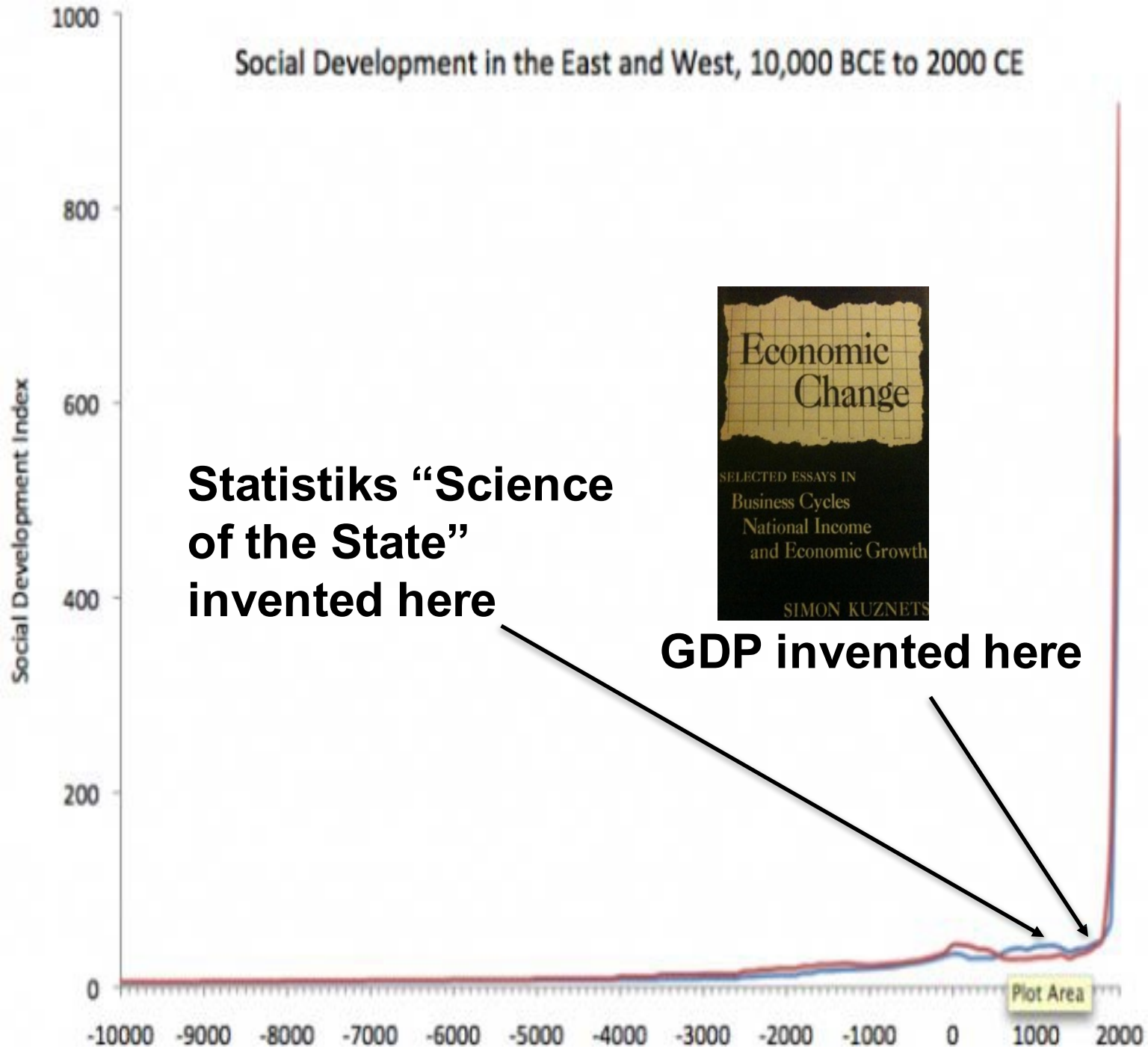
# Why do we measure things? Does it matter? Why or why not? How can it matter more?



MANU.

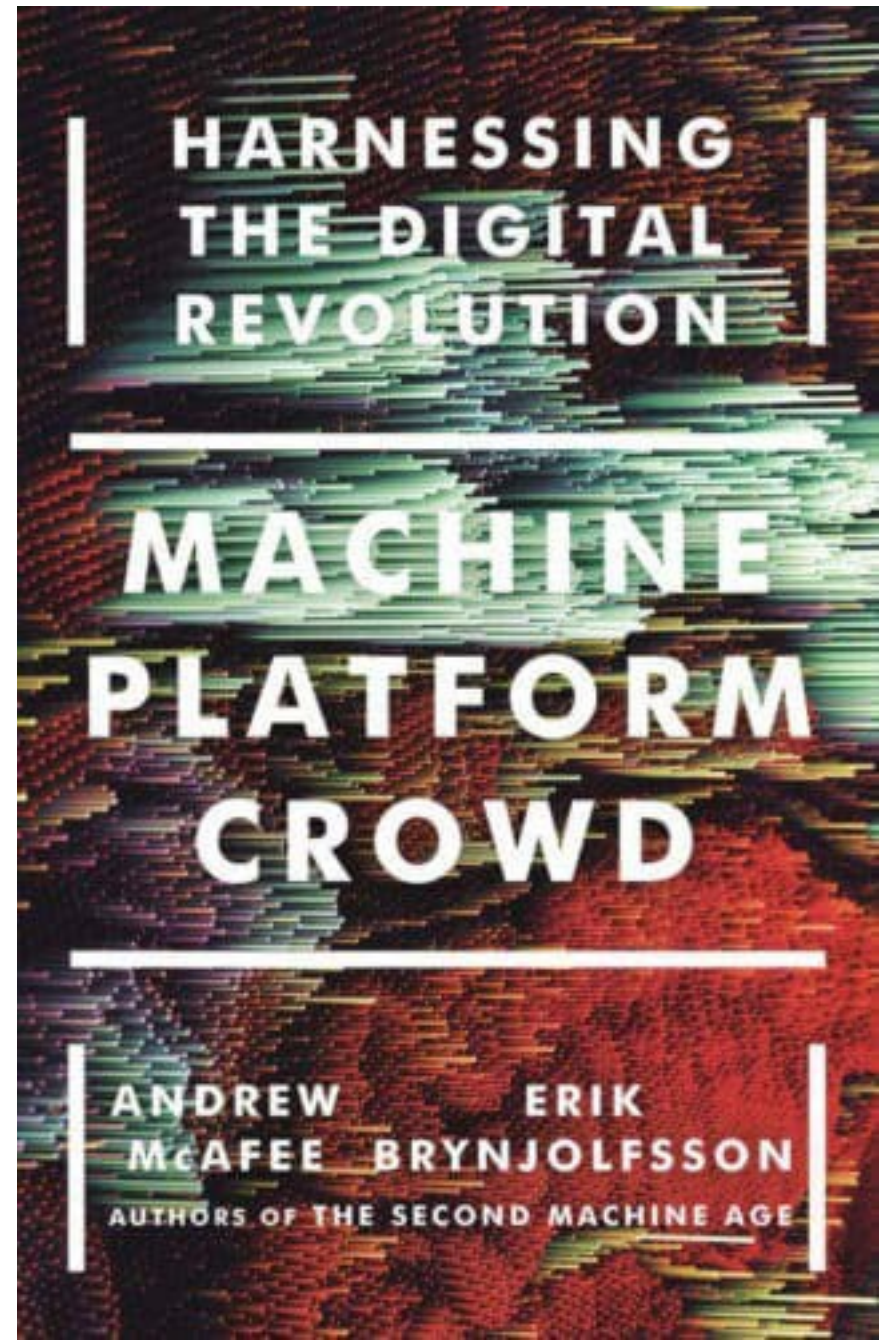
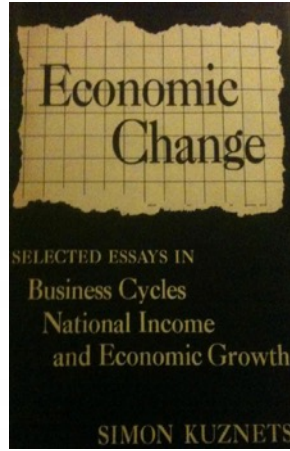


Social Development in the East and West, 10,000 BCE to 2000 CE



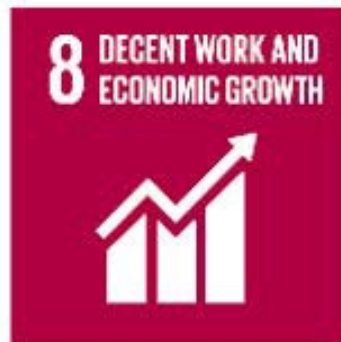
**Statistiks "Science of the State" invented here**

**GDP invented here**





# What about the SDGs?



# The (Big) Data Revolution, democracy, development and the Sustainable Development Goals

## DATA-POP ALLIANCE WORKING NOTE

Reflections on Big Data & the Sustainable Development Goals: Measuring & Achieving Development Progress in the Big Data Era



1. How can (Big) Data help *monitor* the **SDGs** by “filling data gaps” with more granular & disaggregated data—and *what does measuring and monitoring something do to that something?*
2. How can (Big) Data help *promote (or impede?)* the **SDGs** and their underlying human development vision and objectives—including *towards and through lower (or higher?) inequalities?*



# Big Data and AI's relevance for SDG monitoring (from 2015)

## Annex: Uses of Big Data for SDG monitoring

SDGs adopted by the OWG	Big data examples	What is monitored	How is monitored	Country(ies)	Year	Advantages of using big data
1. Poverty eradication	Satellite data to estimate poverty <sup>xiii</sup>	Poverty	Satellite images, night-lights	Global map	2009	International comparable data, which can be updated more frequently
	Estimating poverty maps with cell-phone records <sup>ix</sup>	Poverty	Cell phone records	Cote d'Ivoire	2013-4	
	Internet-based data to estimate consumer price index and poverty rates <sup>x</sup>	Price indexes	Online prices at retailers websites	Argentina	2013	Cheaper data available at higher frequencies
	Cell-phone records to predict socio-economic levels <sup>xi</sup>	Socio-economic levels	Cell phone records	"Major city in Latin America" (Actually Mexico-City)	2011	Data available more regularly and cheaper than official data; informal economy better reflected
2. End hunger, achieve food security and improved nutrition, and promote sustainable agriculture	Mining Indonesian Tweets to understand food price crises <sup>xii</sup>	Food price crises	Tweets	Indonesia	2014	
	Uses indicators derived from mobile phone data as a proxy for food security indicators <sup>xiii</sup>	Food security	Cell phone data and airtime credit purchases	A country in Central Africa	2014	
	Use of remote-sensing data for drought assessment and monitoring	Drought	Remote sensing	Afghanistan, India, Pakistan <sup>xiv</sup>	2004	
3. Health				China <sup>xv</sup>	2008	
	Internet-based data to identify influenza breakouts <sup>xvi</sup>	Influenza	Google search queries	US	2009	Real-time data; captures disease cases not officially recorded; data available earlier than official data
	Data from online searches to monitor influenza epidemics <sup>xvii</sup>	Influenza	Online searches data	China	2013	
	Detecting influenza epidemics using twitter <sup>xviii</sup>	Influenza	Twitter	Japan	2011	
	Monitoring influenza outbreaks using twitter <sup>xix</sup>	Influenza	Twitter	US	2013	
	Systems to monitor the activity of influenza-like-illness with the aid of volunteers via the internet <sup>xx,xxi</sup>	Influenza	Voluntary reporting through the internet	Belgium, Italy, Netherlands, Portugal, United Kingdom, United States	ongoing	
Cell-phone data to model malaria spread <sup>xxii</sup>	Malaria	Cell-phone data	Kenya	2012		

DATA-POP ALLIANCE  
WORKING NOTE

Reflections on Big Data & the Sustainable Development Goals: Measuring & Achieving Development Progress in the Big Data Era

INPUT TO THE BIG DATA AND SDGS CHAPTER OF THE 2015 GLOBAL SUSTAINABLE DEVELOPMENT REPORT

February 2015





A close-up portrait of Sandy Pentland, an older man with white hair and a beard, wearing glasses. The image is dark and serves as a background for the text.

SANDY PENTLAND

**“...Big Data comes from things like location data off of your cell phone or credit card,**

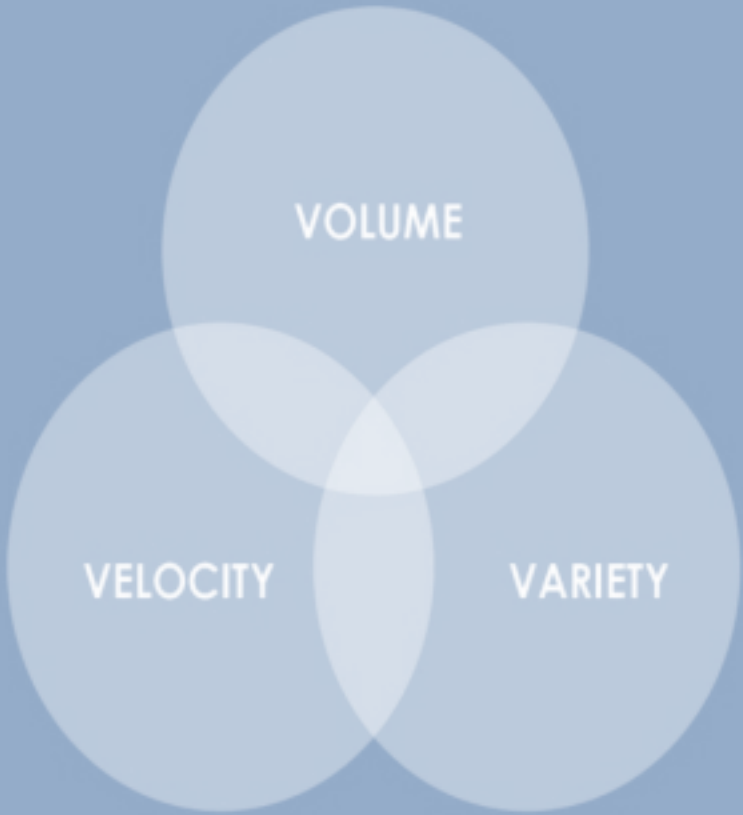
**...the little data breadcrumbs that you leave behind you as you move around in the world...**

REINVENTING SOCIETY IN THE WAKE OF BIG DATA 8.30.12



# Concepts: from big data to Big Data

## EVOLUTION OF THE DEFINITION OF BIG DATA



*circa 2010: the 3 V's of Big Data*

Big Data is Not About the Data!

Gary King<sup>1</sup>

MIT Media Lab  
@medialab

Following

“Big data [is] an ecosystem,” says @ManuLetouze of @datapopalliance, a global coalition that includes the @medialab

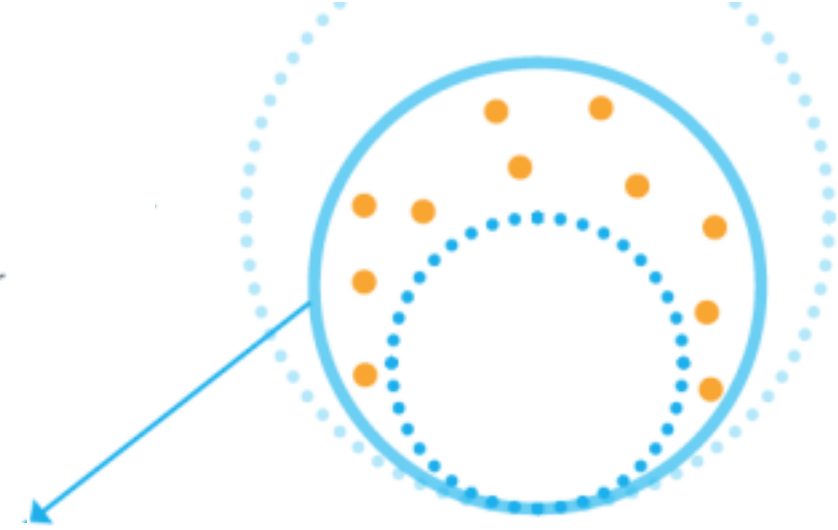


*now: the 3 C's of Big Data*

The **C of crumbs**—i.e. those “digital bread crumbs” or those “digital translations of human actions and interactions passively emitted and captured by digital devices”. At the center of our information societies is the production of massive amounts of data through connected platforms, social networks, and machines. This feature is important as it presides over a fundamental qualitative shift as much as a quantitative one and gives Big Data its deeply political nature.



The **C of capacities**—i.e. tools and methods to collect, aggregate and analyze data. Algorithms—to be defined and discussed below—fall squarely under capacities, and stand firmly at the center of this ecosystem, as both products and drivers of its expansion. Parallel computing is another key aspect without which Big Data would not exist as a techno-social phenomenon as it allows making computations in a fraction of the time—sometimes years—it would take to run them on one machine.



NETWORK  
ANALYSIS  
AND SOCIAL  
PHYSICS

MACHINE  
LEARNING  
TECHNIQUES  
AND ARTIFICIAL  
INTELLIGENCE

TEXT MINING  
TOOLS  
(SENTIMENT  
ANALYSIS, TOPIC  
MODELS, ETC)

SPATIAL  
ANALYSIS AND  
GEOGRAPHIC  
INFORMATION  
SYSTEMS (GIS)

CROWDSOURCING

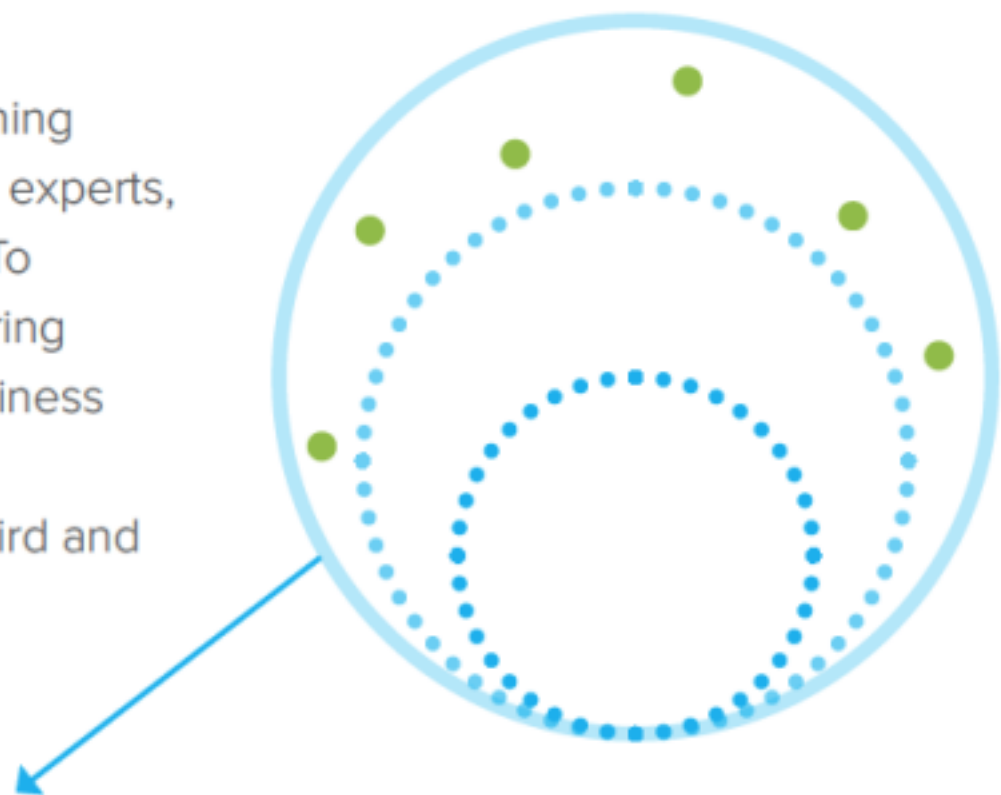
DATASET  
DISAGGREGATION  
AND UNIFICATION

DISTRIBUTED  
NETWORKS OF  
DEVICES

CLOUD  
COMPUTING  
AND STORAGE



The **C of communities**—i.e. all those involved in generating, governing and using data, including data producers, end users, policymakers, experts, privacy advocates and civic hacker communities. Namely, groups. To date the two constituencies that have been the most active in leveraging algorithms to make decisions of not as the centerpiece of their business are large private companies and government agencies— notably those in charge of surveillance activities—with academia coming third and organized advocacy groups and networks (e.g. in the humanitarian space) coming fourth.



ACADEMIA AND  
RESEARCH

CIVIL SOCIETY/  
NGOs

MULTILATERAL  
INSTITUTIONS

GOVERNMENT/  
NATIONAL  
STATISTICAL  
OFFICES (NSOs)

PRIVATE SECTOR  
(e.g. TELECOM,  
FINANCIAL, etc)

# Concepts: from big data to Big Data

## **1. Big data (as *data*):**

*“Digital translations of human actions, interactions and transactions picked up by digital devices and services.”*

## **2. Big Data (as a *field of research and practice*):**

*an ecosystem of the 3 Cs of Big Data as data ‘crumbs’, capacities (human and technical), and communities ] producing and leveraging information to shape decisions.*

# Functions of Big Data

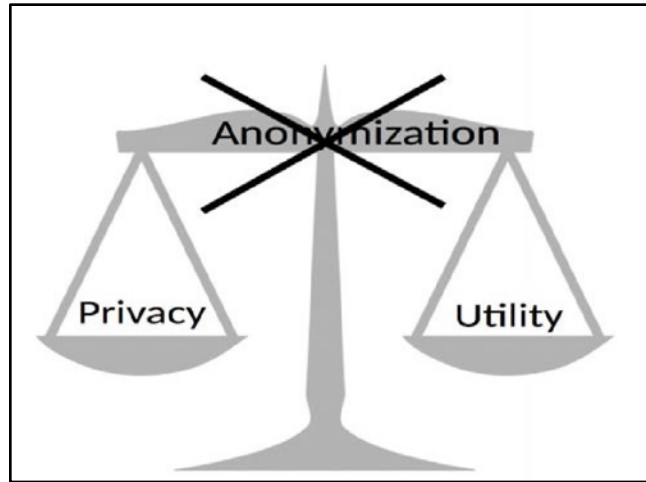
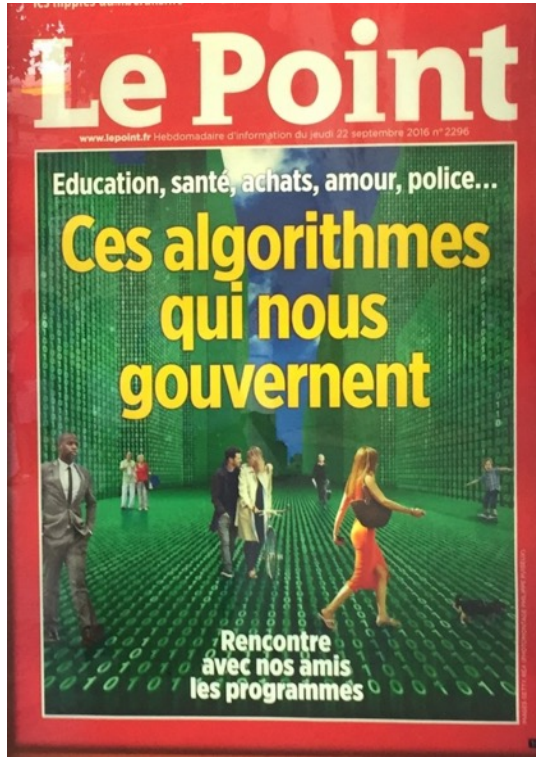
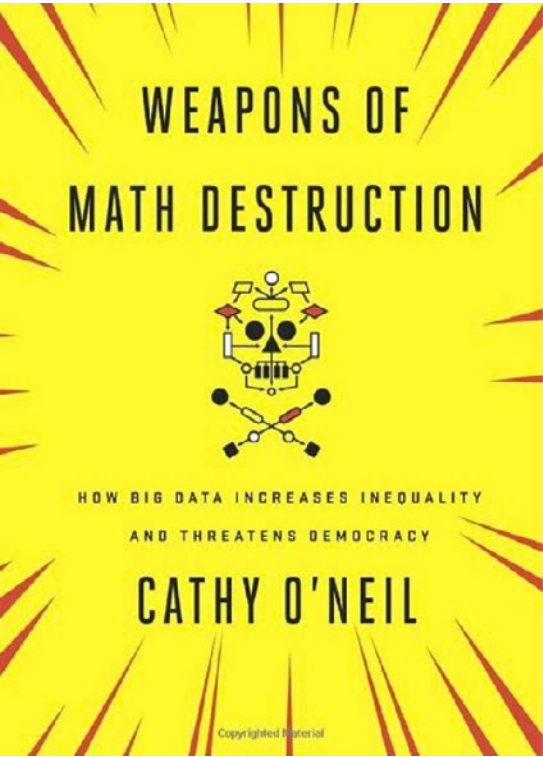
1. Descriptive
2. Predictive
  - i. Forecasting
  - ii. Nowcasting
3. Prescriptive
4. Discursive





# Is this all really bad?

*Data centers CO<sub>2</sub> emissions > entire airline industry*



The real digital divide is between families that limit screen time and those that don't

## Computational Privacy

How human behavior bounds privacy and what we can do about it

Yves-Alexandre de Montjoye  
Twitter: yvesalexandre  
MIT Media Lab – Human Dynamics

Is data a danger to the developing world?

By Kate Crawford

Nov 2 2015

f 128 t 82 in 1

Part 1:  
**Genesis, Context, Concepts  
and Questions of the 4<sup>th</sup> Industrial  
'Data' Revolution**

Part 2:  
**Statistical Measurement and  
Sustainable Development in the  
Age and Big Data and AI**

Part 3:  
**Pillars and Pathways of a People-  
Centered, Data-Enabled Human  
Development Revolution:  
Towards and Human AI**



# an illustrated introduction to Predicting socioeconomic levels through cell-phone data

## Question:

so, how is it possible to predict an area's socioeconomic - or poverty- level from the cell - phone data it emits?

## step ①



## step ②



"these digital traces, recorded by every telecom operator, are <u>Call Detail Records</u> or CDRs, metadata that look like that"

CALLER ID	CALLER LOCATION	RECIPIENT ID	RECIPIENT LOCATION	CALL TIME	CALL DURATION
X36872	2°24'22"	A8C492	5°38'49"	2014.04.01	01.12.27
9748Y	35°49'58"	TC73646	31°12'22"	ET 17 22	

## TELECOM OPERATOR DATA CENTER



"and these CDRs will show differences in calling patterns between different areas ..."





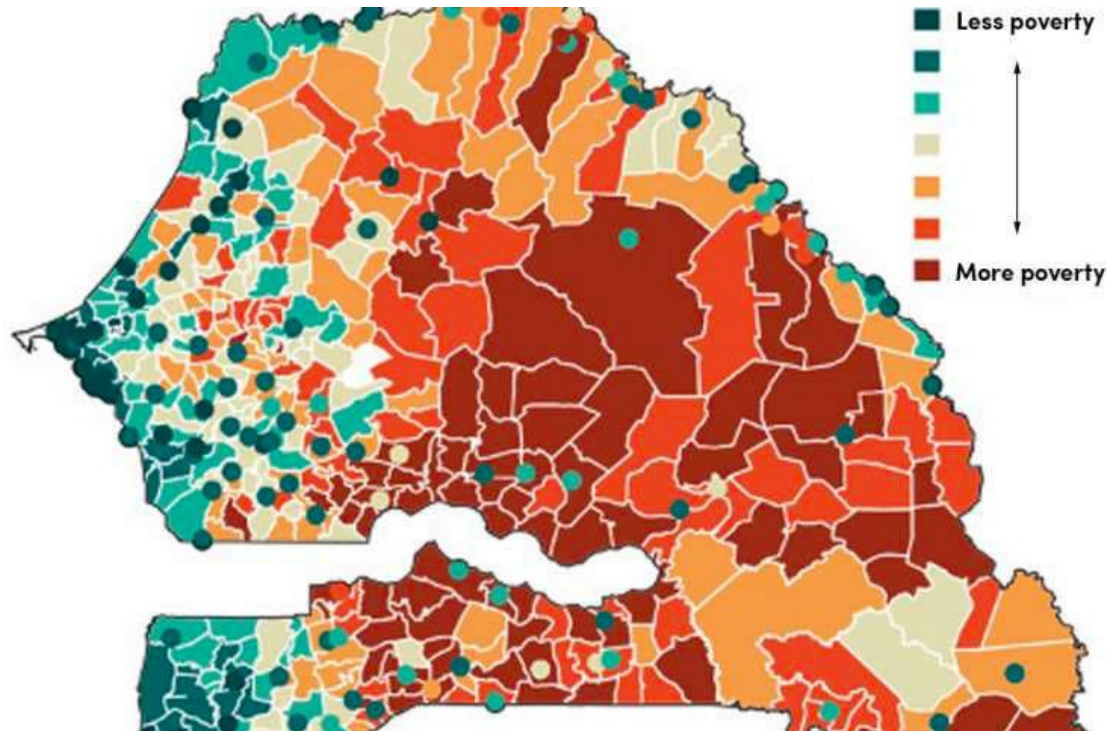


# Scientific Prize and Ethics Mention: Construction of socio-demographic indicators with digital breadcrumbs

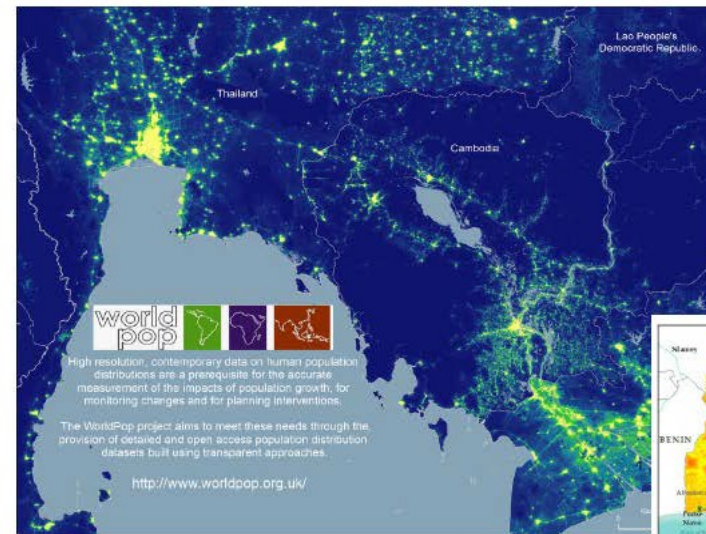
F. Bruckschen <sup>(1)</sup>, T. Schmid <sup>(2)</sup>, T. Zbiranski <sup>(1)</sup>

We show that socio-demographic indicators such as population, age, literacy, poverty, religion, ethnicity, electricity supply and others can be estimated in unprecedented detail and virtually ad-hoc using antenna-to-antenna traffic data only. We offer a uniform approach that can be easily extended to other variables. Results are tested for spatio-temporal robustness and visualized as heat maps.

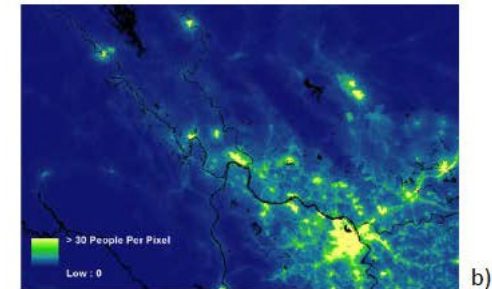
(1) Humboldt Universität Berlin, Germany - (2) Freie Universität Berlin, Germany



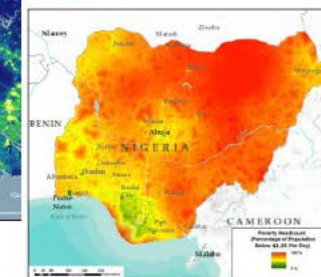
FLOWMINDER.ORG



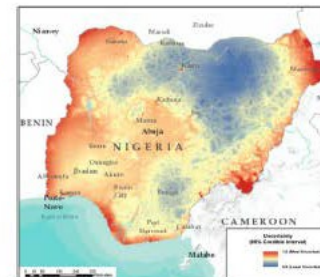
a)



b)



c)



d)

Figure 1: a) WorldPop population density distributions for the the Mekong region; b) close-up picture of population distributions (100x100m) for the Hanoi region; c) Poverty headcounts for Nigeria (<1.25 USD/day) per 1 km<sup>2</sup>; d) Uncertainty in poverty headcount estimates per 1 km<sup>2</sup> area.

# Since 2013-14 : The ‘Data Revolution’ and/at the UN



**2013:** Call for a *“data revolution for sustainable development (...) to improve the quality of statistics and information available to people and governments” ...*

*“Data are the lifeblood of decision-making and the raw material for accountability.*

*Governments, companies, researchers and citizen groups are in a ferment of experimentation, innovation and adaptation to the new world of data, a world in which data are bigger, faster and more detailed than ever before. This is the data revolution.”*





# Measurement ↔ Development?

↑  
Level of Development



## HUMAN DEVELOPMENT INDEX (HDI) RANKS 2014

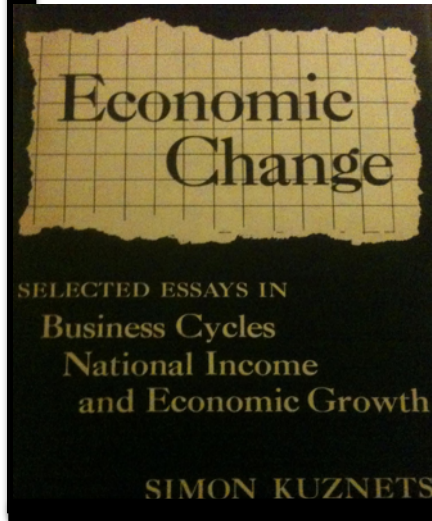
### Top 5 Countries

- 1 Norway
- 2 Australia
- 3 Switzerland
- 4 Netherlands
- 5 United States



### Bottom 5 Countries

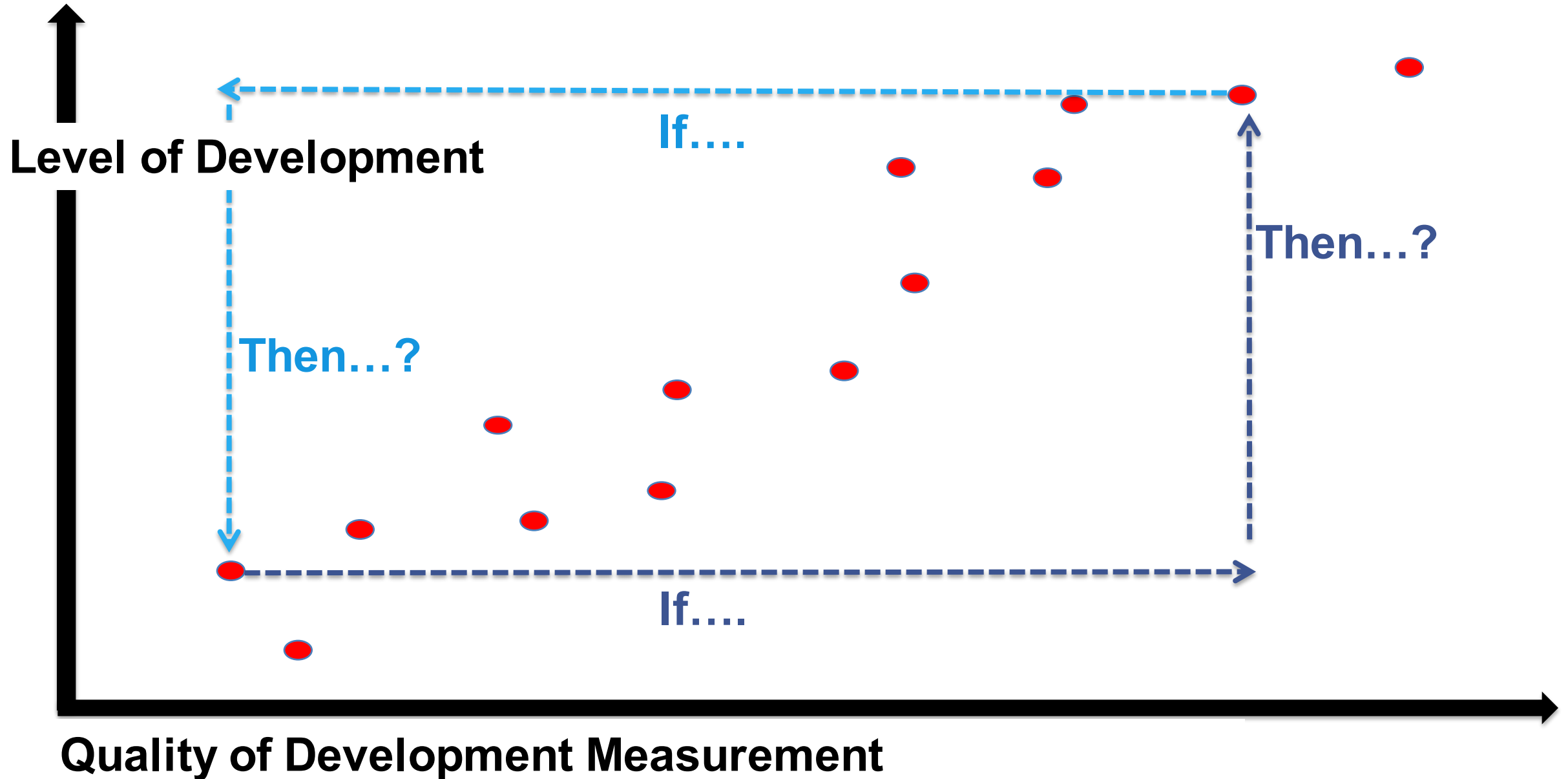
- 1 Sierra Leone
- 2 Chad
- 3 Central African Republic
- 4 DR Congo
- 5 Niger



→  
Quality of Development Measurement



# Better Data → Better Decisions → Better Development?



# Better Data → Better Decisions → Better Development?

MORTEN JERVEN

## POOR NUMBERS

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



*“Despite the many publications I think there are still many holes in our knowledge. There is a further need for empirical research on the lines of ‘political ethnography of indicators’.*

***Particularly is there a gap in theory and empirical studies on the line of causality from ‘data’ to ‘decisions’”***

*Morten Jerven, 2015*

# Who wants to know?

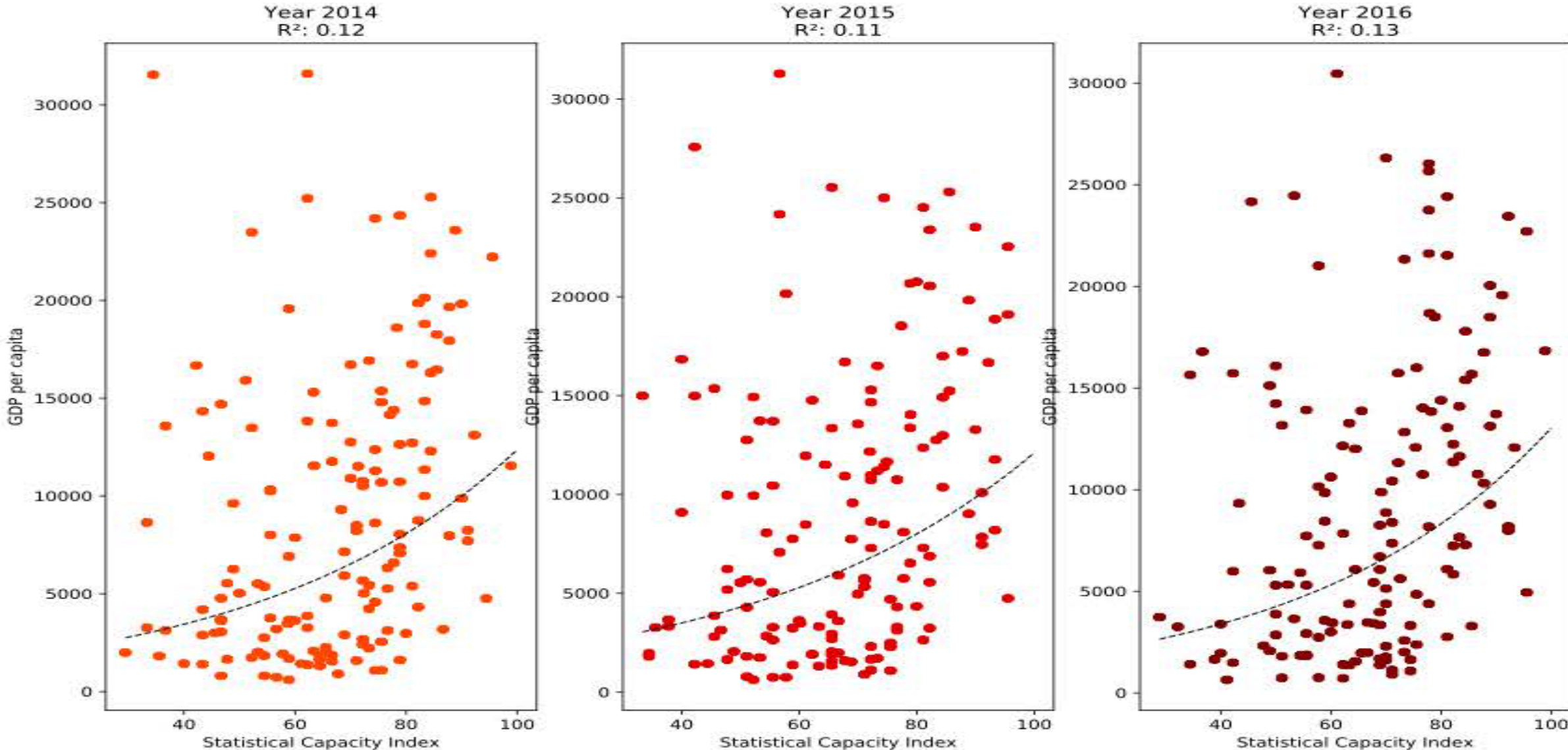
The Political Economy  
of Statistical Capacity  
in Latin America

Eduardo Dargent, Gabriela Lotta,  
José Antonio Mejía, and Gilberto Moncada



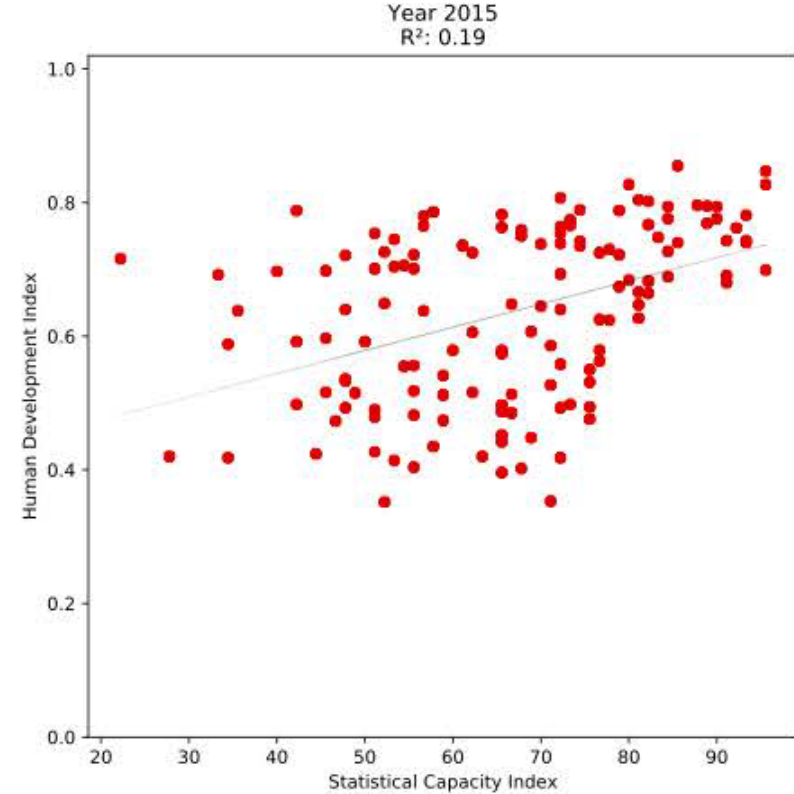
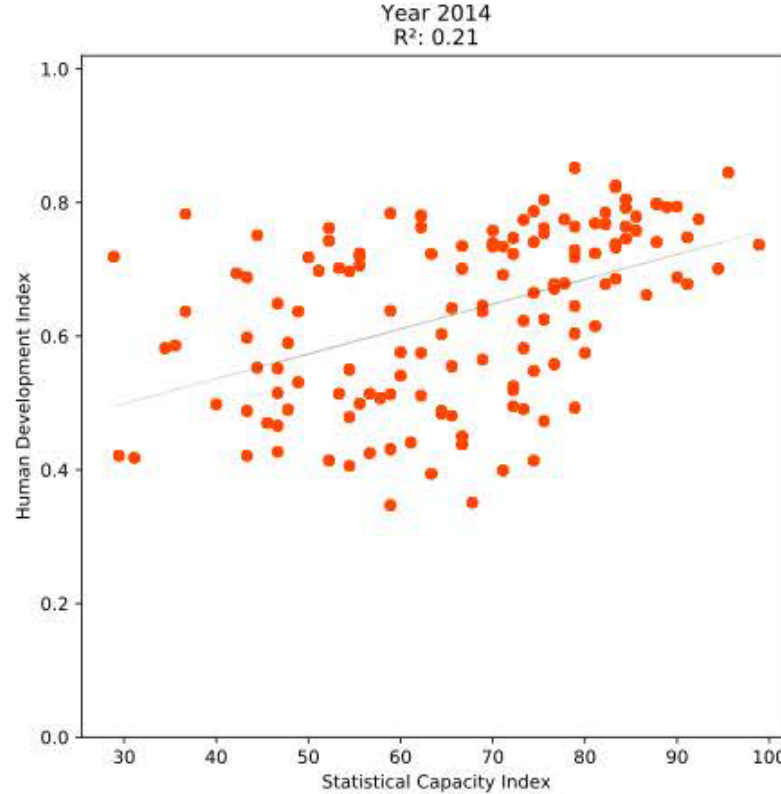
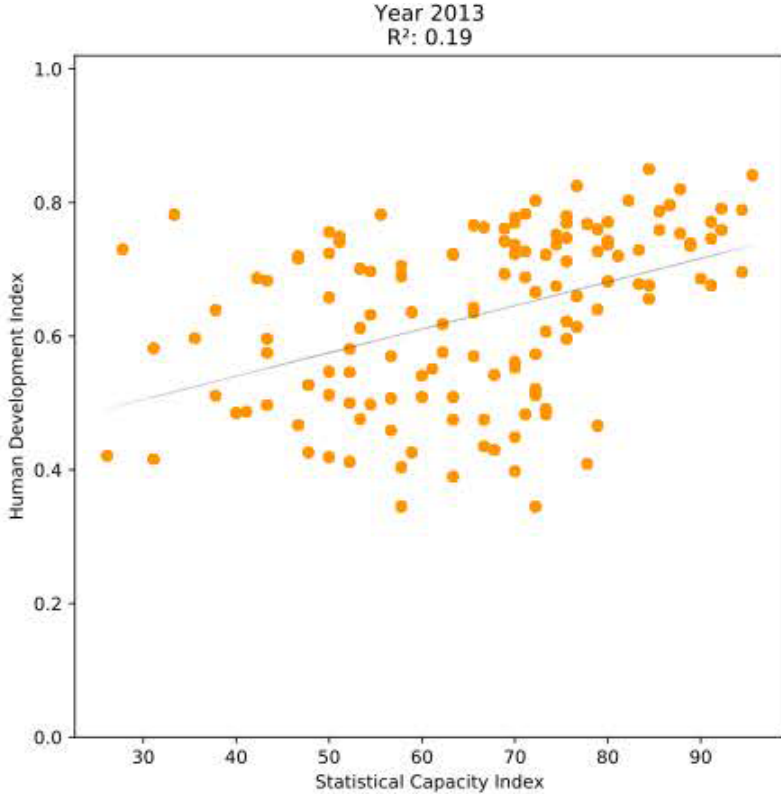
## What does the evidence suggest? Why?

# Development vs. statistical capacity: a rather weak link



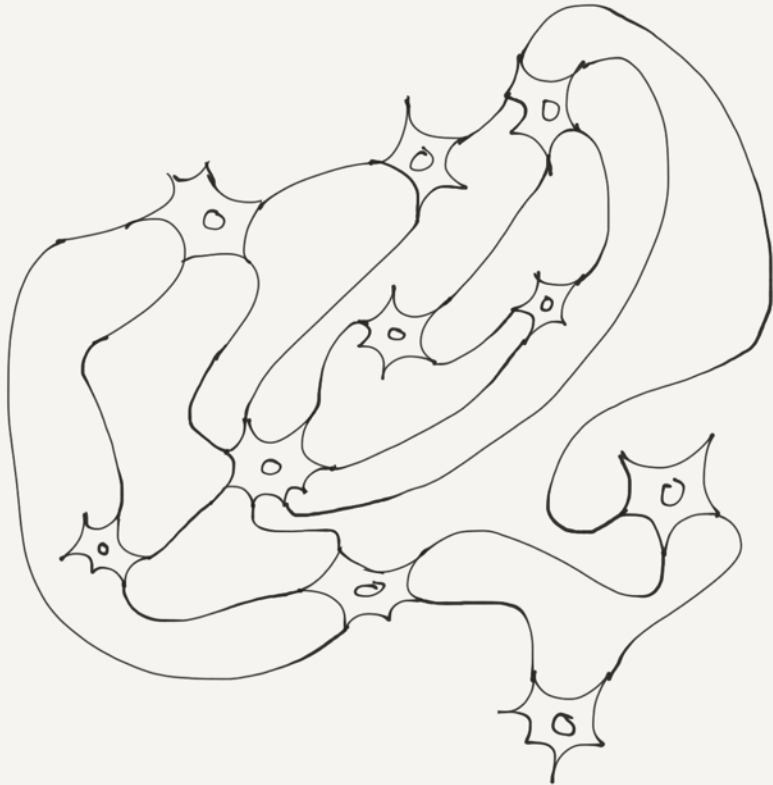


# Development vs. statistical capacity: a rather weak link

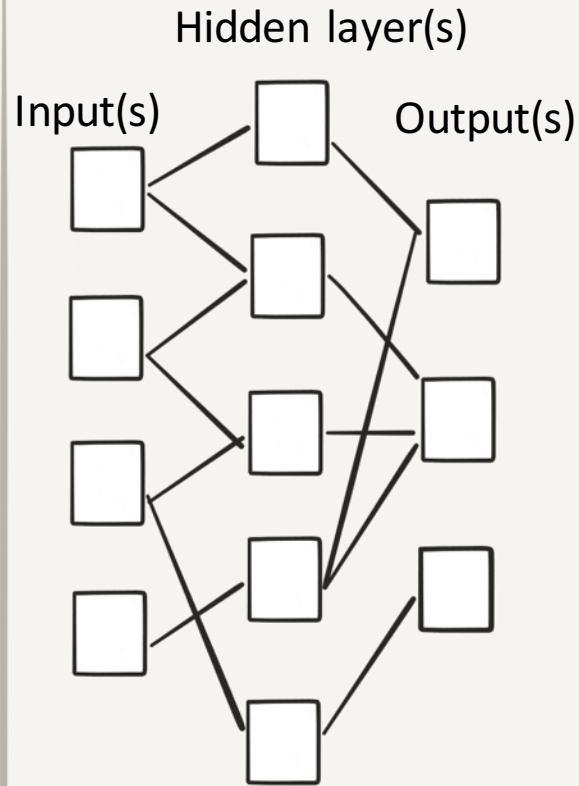


**So, what do we do?**

# The (re)rise of AI.



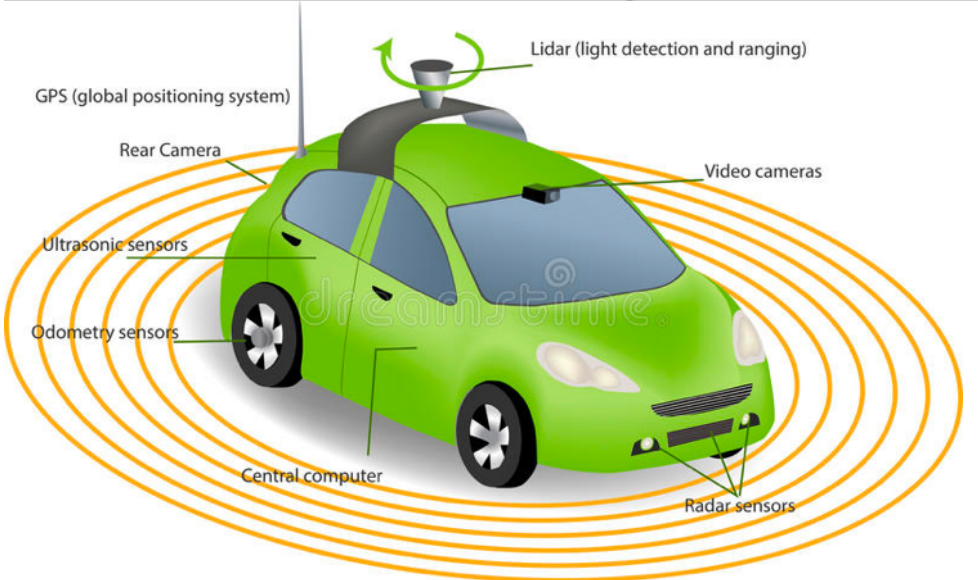
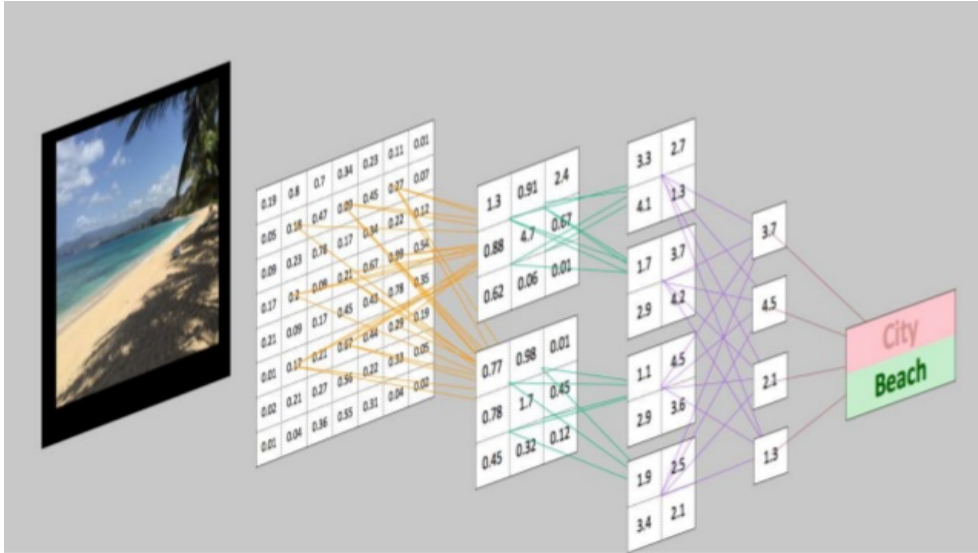
BIOLOGICAL



ARTIFICIAL

Artificial intelligence—broadly-- is the **simulation of human intelligence processes by computer systems, especially artificial neural networks (ANNs)** inspired by the biological neural networks that constitute animal brains, which can "learn" (i.e. progressively improve performance on) through iterations and feedback. **Basically it's algorithms that learn to automate parts or all of tasks, and the machines they power.** (It's also what has not been invented yet)

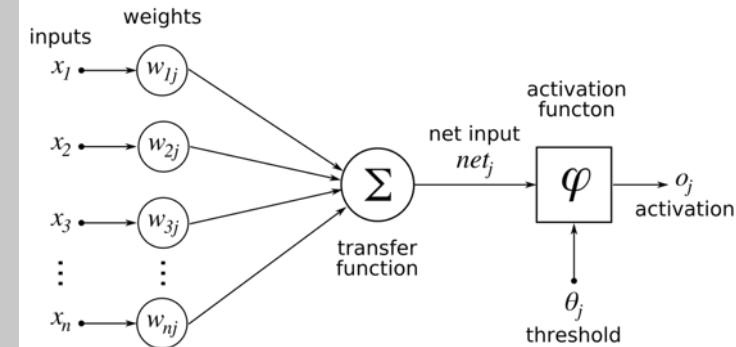
# The basics of AI is learning through *many* feedbacks



1. Try to guess / recognize. Right or Wrong?
  2. Correct: +1. Reward!
  3. Incorrect: -1. Penalty.
  4. Repeat and learn through a feedback loop.
- ➔ (The) machine (is) learning!



# From Big Data to AI: what can we learn?

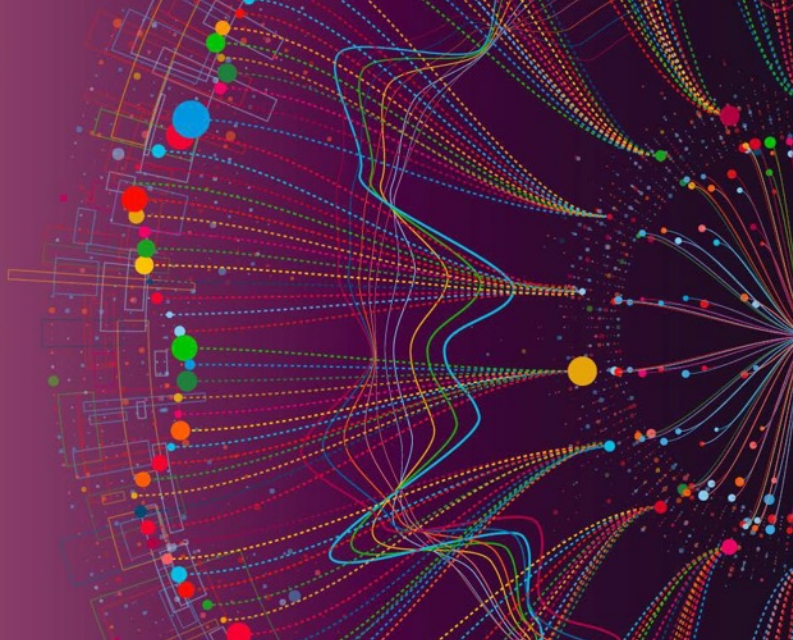


*...it's neither new nor black magic...*

1. It is **at least 60+ years old**.
2. It still generalizes poorly. It has no sense of context. **It is still pretty stupid.**
3. We are **far from general AI**.
4. **Humans are still in control** (for better or worse).

*...but...*

1. The (good) magic / core of the current AI is **the credit assignment function to encourage and reinforce neurons / functions that help the most achieve the goal** (and reverse if not).
2. The key difference and is data. **Big Data.**



ITU Journal: *ICT Discoveries*, Special Issue No. 2, 6 Dec. 2018

## TOWARDS A HUMAN ARTIFICIAL INTELLIGENCE FOR HUMAN DEVELOPMENT

Emmanuel Letouzé<sup>1</sup>, Alex Pentland<sup>2</sup>

<sup>1</sup>Data-Pop Alliance, MIT Media Lab, and OPAL, <sup>2</sup>MIT and Data-Pop Alliance, and OPAL

---

**Abstract** – *This paper discusses the possibility of applying the key principles and tools of current artificial intelligence (AI) to design future human systems in ways that could make them more efficient, fair, responsive, and inclusive.*

**Keywords** – Artificial intelligence, big data, human development, open algorithms, fourth industrial revolution

# A positive vision: Towards “Human AI” ecologies

MIT Prof Alex ‘Sandy’ Pentland:

*“The big question that I’m asking myself these days is **how can we make a human artificial intelligence?** (...) I don’t want to think small—people talk about robots and stuff—I want this to be global. (...)”*

*What would happen if you had a **network of people where you could reinforce the ones that were helping and maybe discourage the ones that weren’t? That begins to sound like a society or a company**”.*

The Human Strategy. [www.thehumanstrategy.mit.edu](http://www.thehumanstrategy.mit.edu)



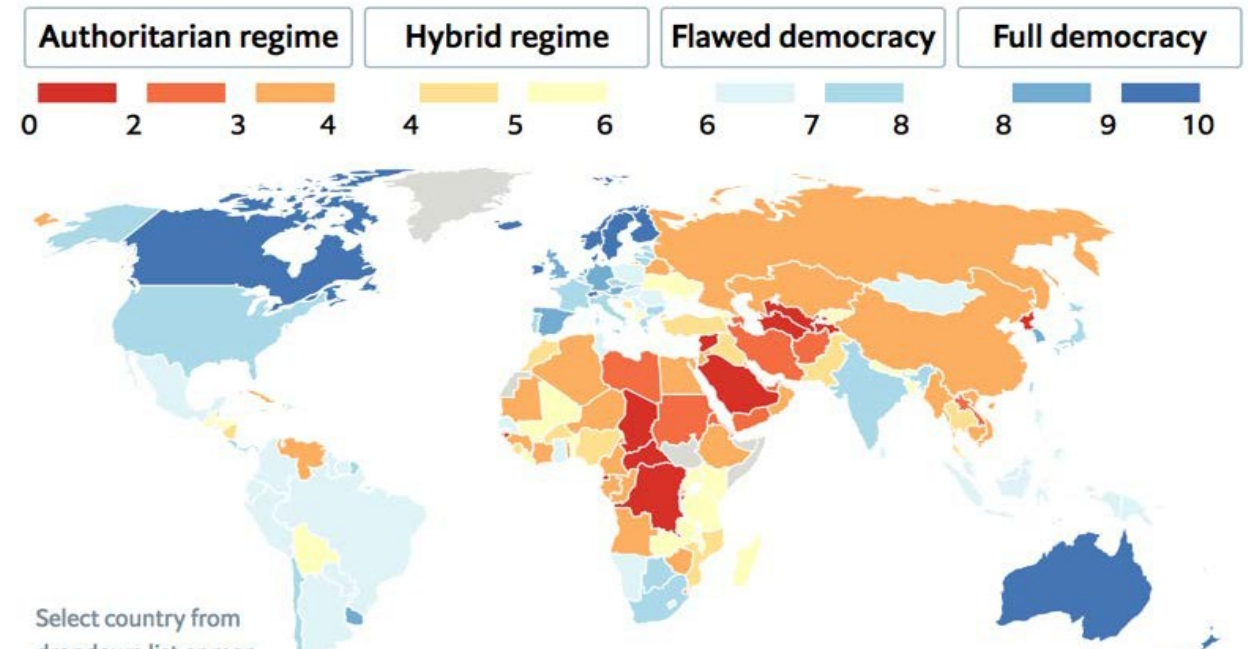


# Main challenges to a Human AI

1. Powerful agents have an **incentive for this not to work** (e.g. economic and political elites benefit from status quo).
2. Most **societies / countries currently lack appropriate data connections, capacities, and culture** for this.
3. There is **widespread digital and analog distrust, disdain, echo chambers, alternative facts narratives**, hampering cooperation, consensus, compromise.

**It is very hard for facts and measures to “matter”**

**The Economist Intelligence Unit's Democracy Index**  
167 countries scored on a scale of 0 to 10 based on 60 indicators

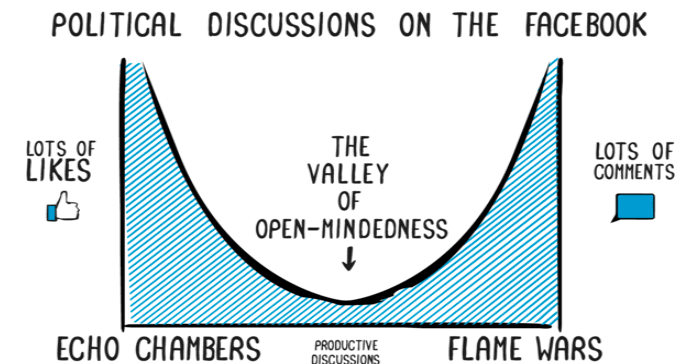


THE NEW YORKER

## WHY FACTS DON'T CHANGE OUR MINDS

*New discoveries about the human mind show the limitations of reason.*

By Elizabeth Kolbert



# NERVOUS STATES

DEMOCRACY  
AND  
THE DECLINE  
OF REASON

WILLIAM  
DAVIES

“In this age of intense political conflict, we sense objective fact is growing less important. Experts are attacked as partisan, **statistics and scientific findings are described as propaganda**, and public debate devolves into personal assault. How did we get there and what can we do about it?”

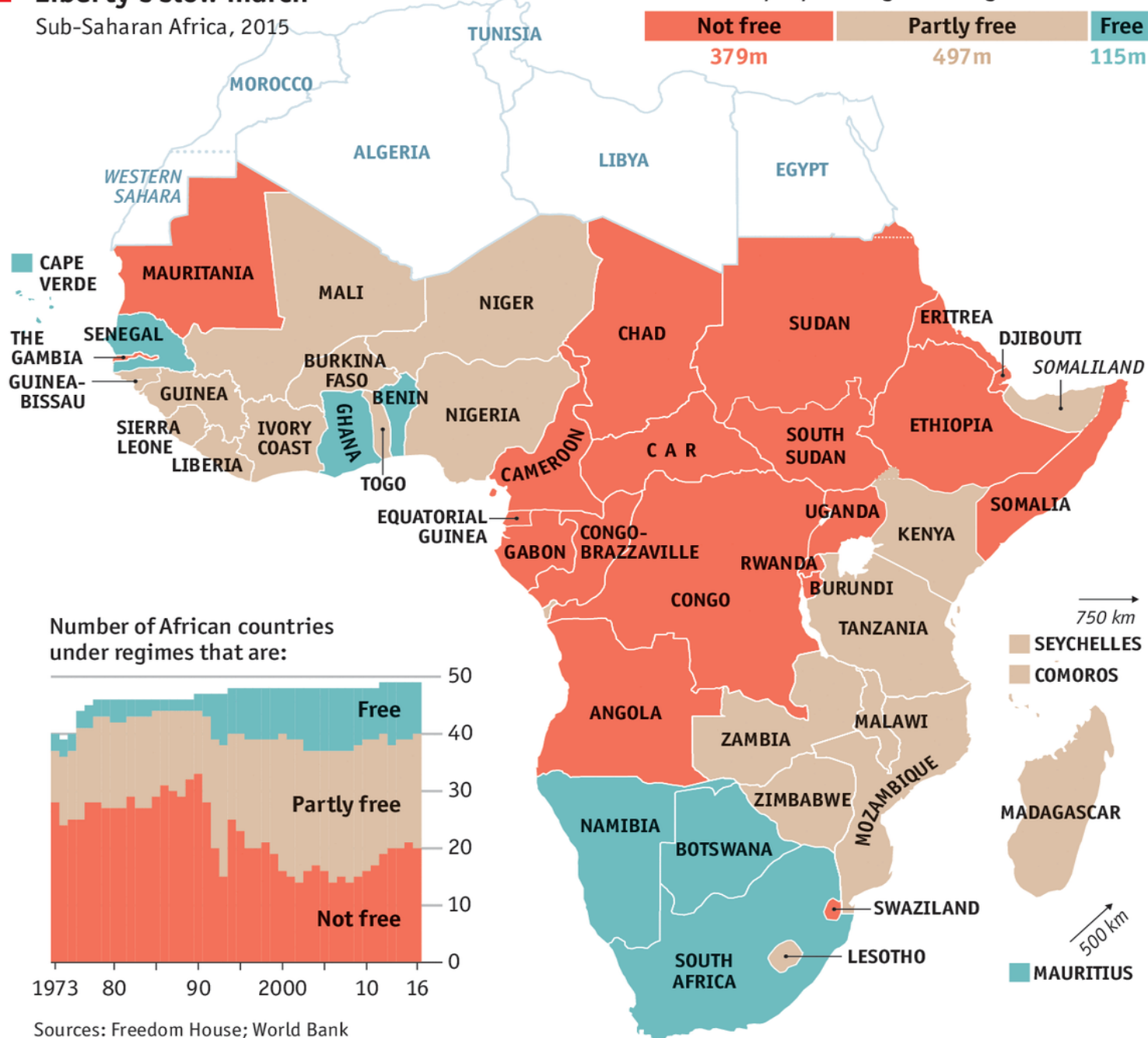
**MLTalks: How Data Killed Facts**  
Jill Lepore in conversation with  
Andrew Lippman



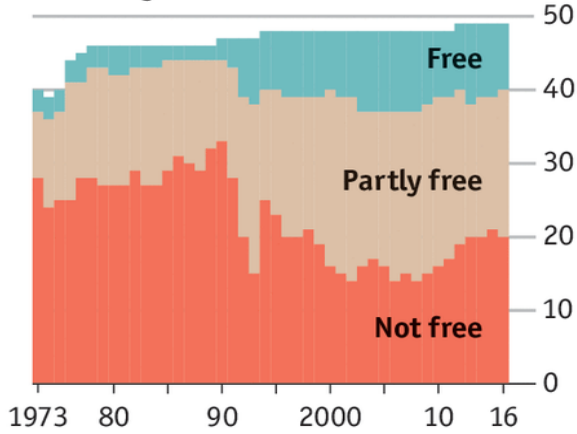
# Liberty's slow march

Sub-Saharan Africa, 2015

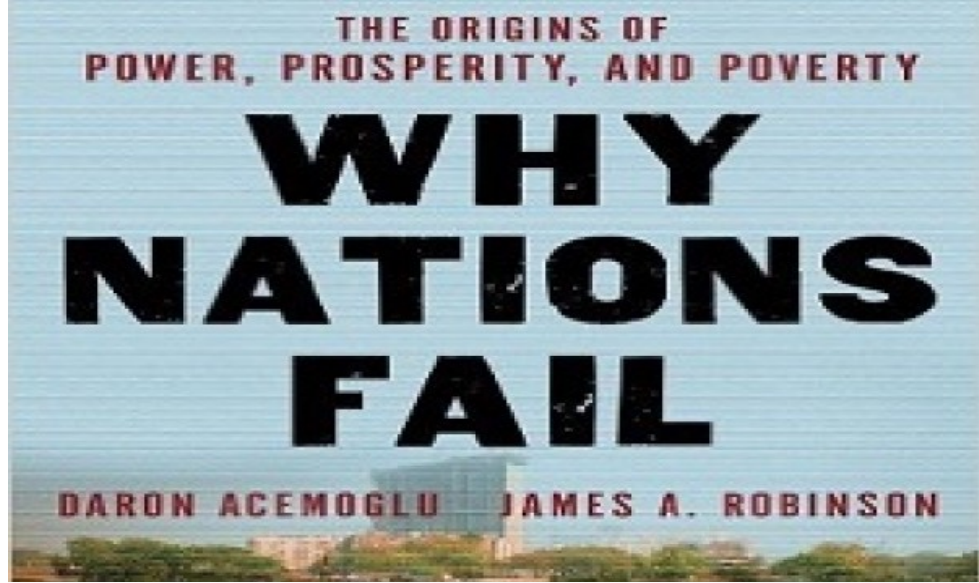
Number of people living under regimes that are:



Number of African countries under regimes that are:



Sources: Freedom House; World Bank  
Economist.com



Norway topped the ranking and North Korea came last.

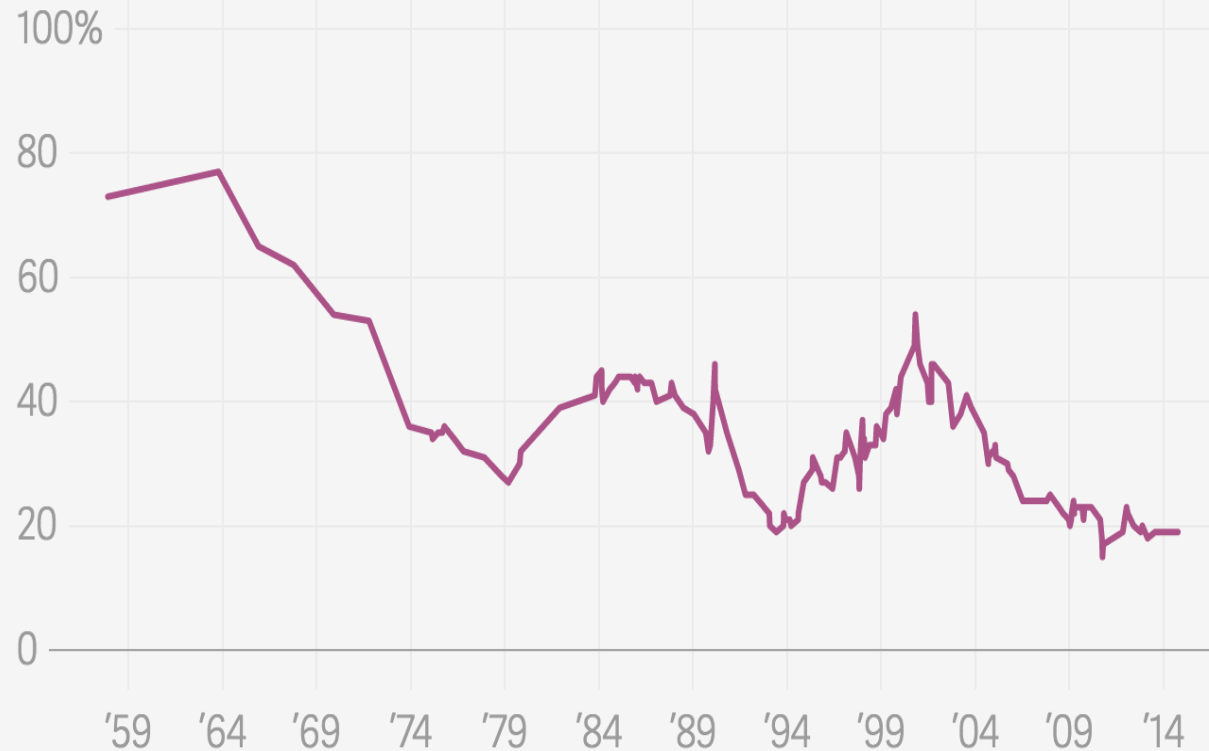
Top 10	Bottom 10
1. Norway	158. Uzbekistan
2. Iceland	T-159. DR Congo
3. Sweden	T-159. Saudi Arabia
4. New Zealand	161. Tajikistan
5. Denmark	162. Turkmenistan
T-6. Canada	163. Equatorial Guinea
T-6. Ireland	164. Central African Rep.
8. Switzerland	165. Chad
9. Finland	166. Syria
10. Australia	167. North Korea

“Democracy is in trouble in the West, in the mature democracies of western Europe and the US, which are no longer obvious beacons for those striving for democracy in the nondemocratic world,” the EIU said.



# The US has been downgraded to a “flawed democracy,” but not just because of Trump

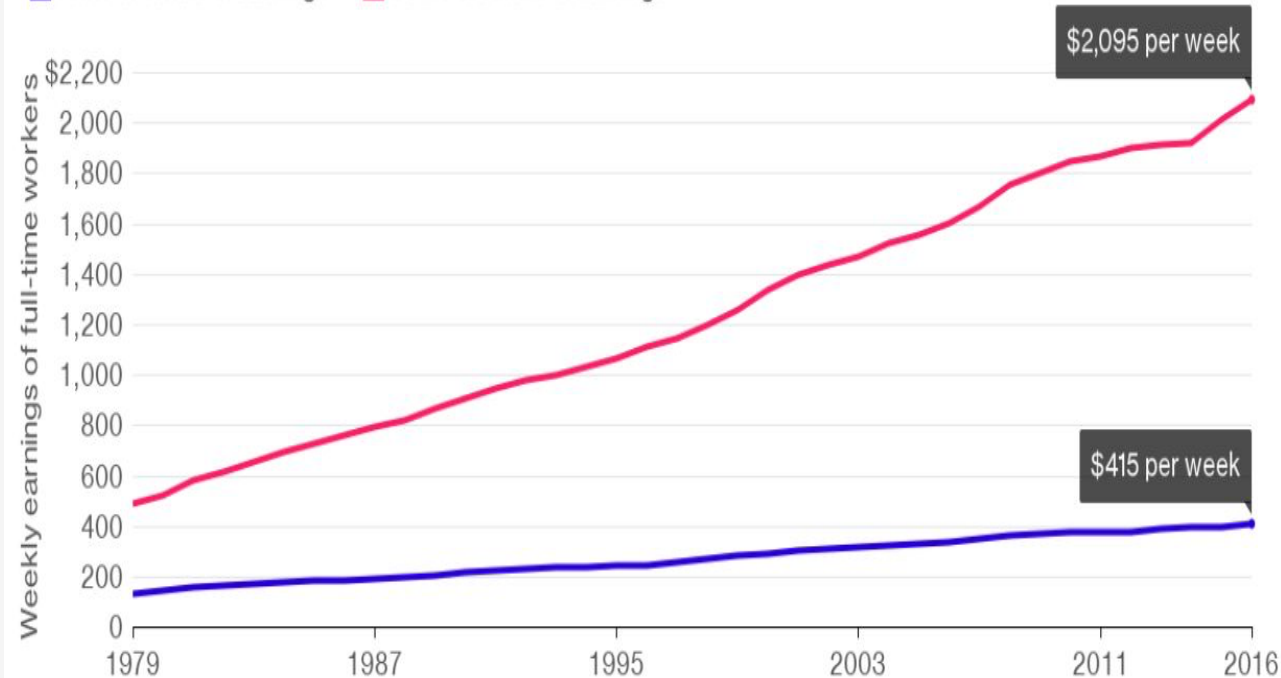
US public trust in government



## The Rich Get Richer

Gap between top and bottom wage earners continues to widen

■ 10th Percentile of Earnings ■ 90th Percentile of Earnings



# Data, Statistics, Measurement, Development, and Democracy

## Statistics 2.0 The next level

By Enrico Giovannini

(2010)

**Official statistics' two main functions:**

1) **provide society with “knowledge of itself, on which to base its own choices and evaluate the effects of political decisions.”.**

2) **The second function of official statistics is provides a deliberative space where what is worth measuring, how it is measured, and for which purpose it is measured is freely and openly debated—to act as “a debated public institution”.**

Part 1:  
**Genesis, Context, Concepts  
and Questions of the 4<sup>th</sup> Industrial  
'Data' Revolution**

Part 2:  
Statistical Measurement and  
Sustainable Development in the Age  
and Big Data and AI

Part 3:  
**Pillars and Pathways of a People-  
Centered, Data-Enabled Human  
Development Revolution:  
Towards and Human AI**



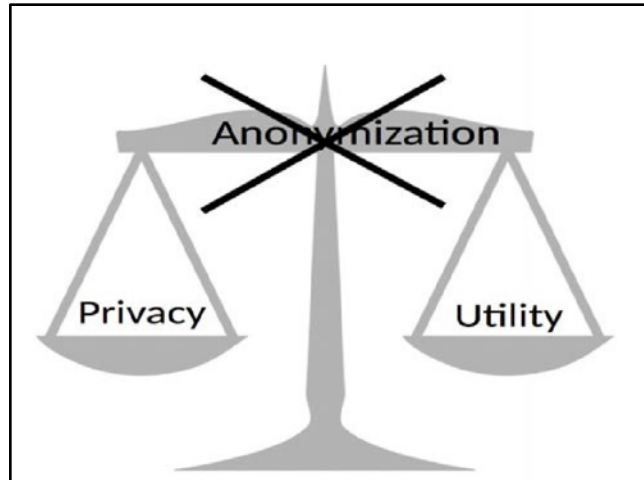
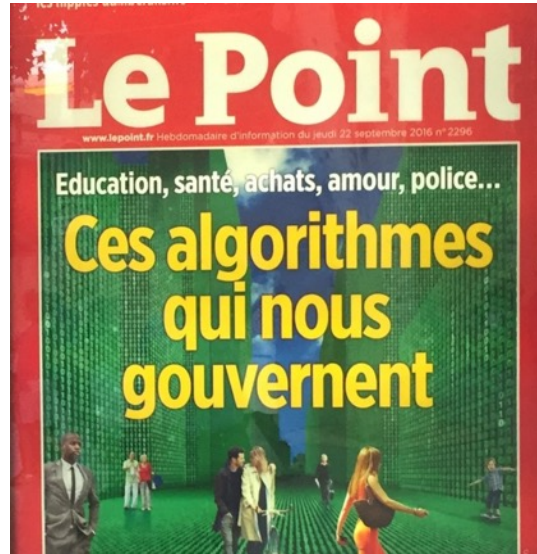
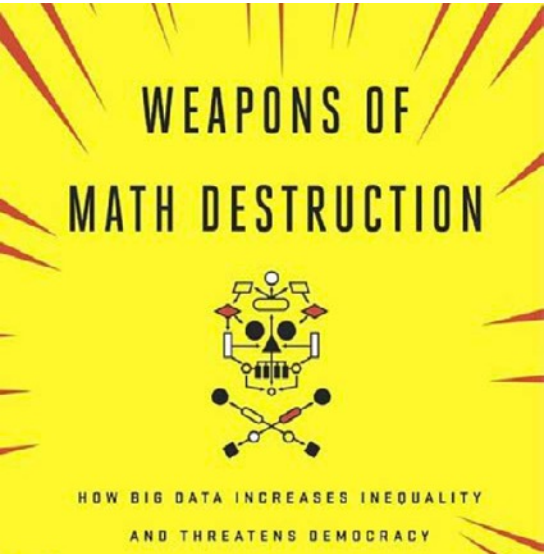
# There are reasons to be concerned...

*Data centers CO<sub>2</sub> emissions > entire airline industry*

# IS DEMOCRACY DYING?

A WARNING FROM EUROPE:  
The Worst Is Yet to Come  
BY ANNE APPLEBAUM

HOW AI COULD GIVE RISE TO TYRANNY  
by Yuval Noah Harari



## MLTalks: How Data Killed Facts

Jill Lepore in conversation with Andrew Lippman

The real digital divide is between families that limit screen time and those that don't

### Computational Privacy

How human behavior bounds privacy and what we can do about it

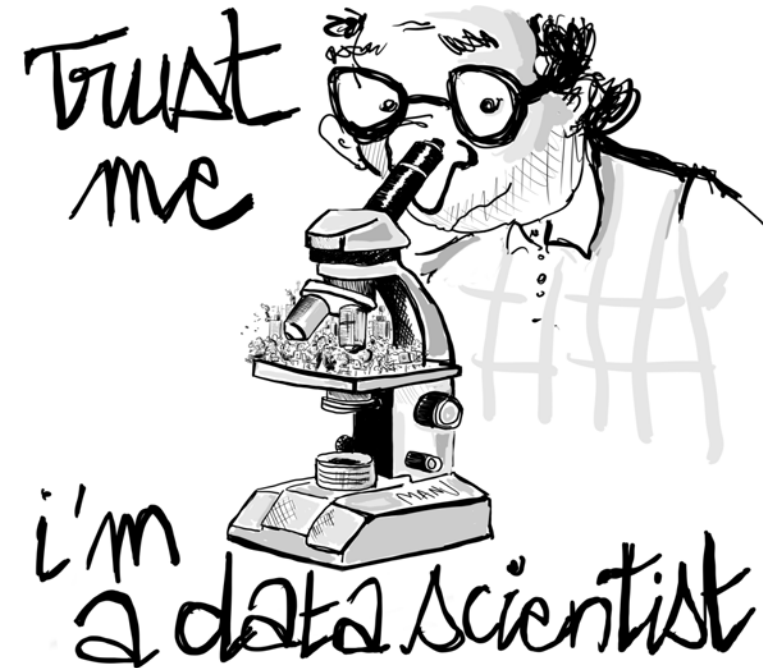
Yves-Alexandre de Montjoye  
Twitter: yvesalexandre  
MIT Media Lab – Human Dynamics

### Is data a danger to the developing world?

By Kate Crawford

Nov 2 2015

f 128 t 82 in 1



# But we can't give up on (data) science and technology

By Alex "Sandy" Pentland

# Saving Big Data from Itself

*Journal of Economic Perspectives—Volume 17, Number 4—Fall 2003—Pages 167–190*

## The Demographic Transition: Three Centuries of Fundamental Change

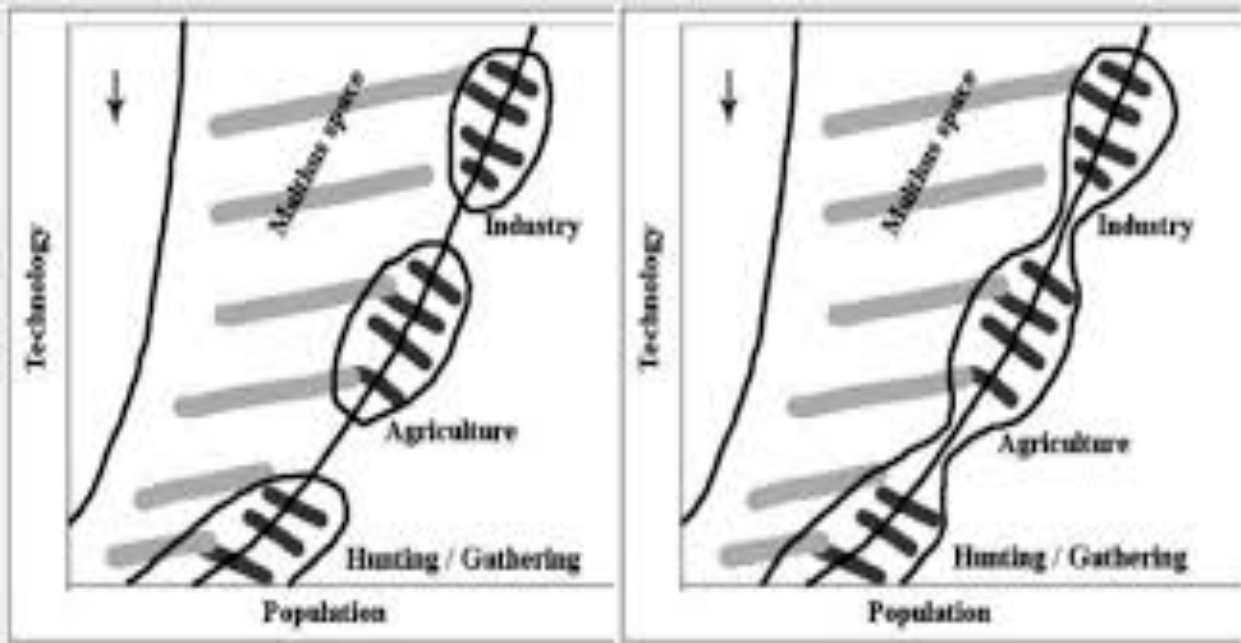
Ronald Lee

*Before the start of the demographic transition, life was short, births were many, growth was slow and the population was young.*

### Saving Big Data from Big Mouths

Those who would condemn big data ought to try making something

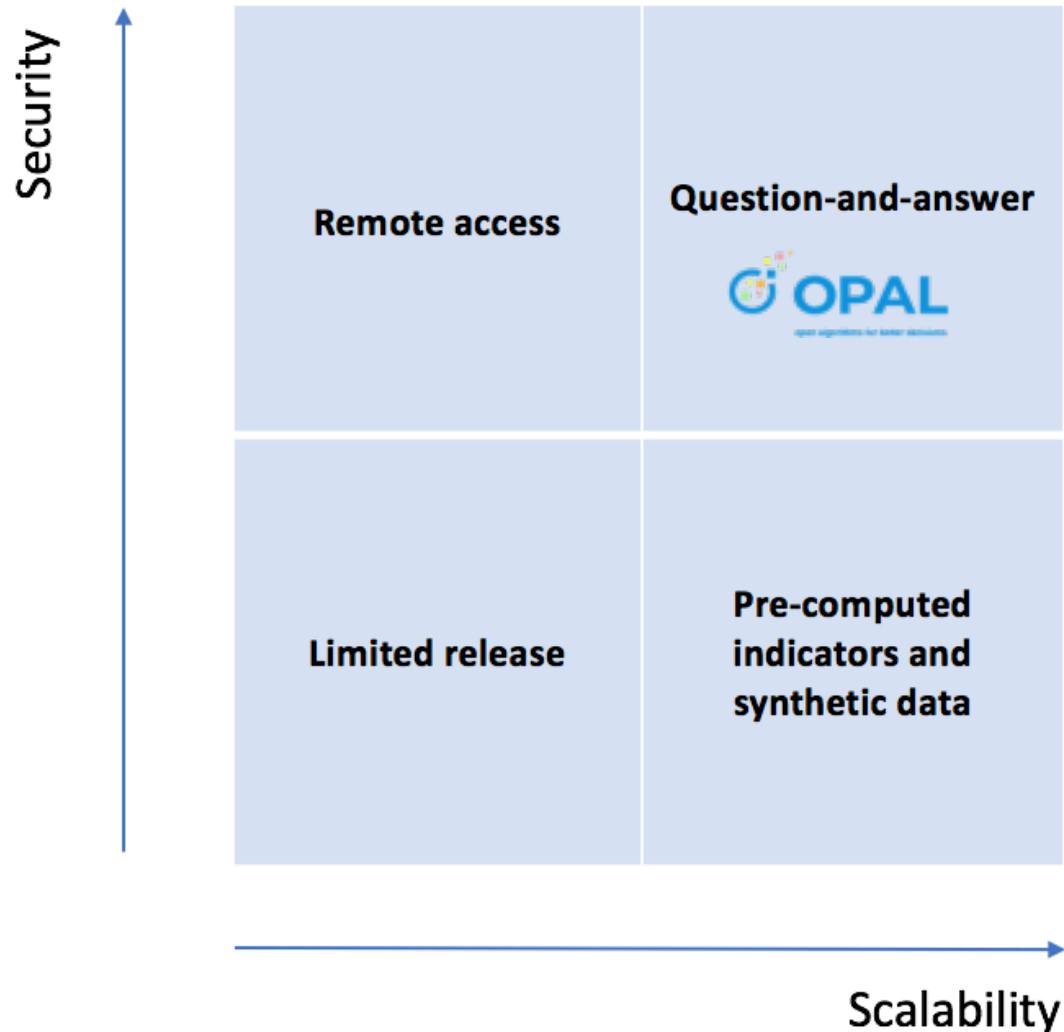
By Cesar A. Hidalgo | April 29, 2014



Ronald D. Lee: "Malthus and Boserup: A dynamic synthesis (1984)



# How to 'open' private sector data safely, ethically, at scale?



## On the privacy-conscious use of mobile phone data

Yves-Alexandre de Montjoye , Sébastien Gambs, Vincent Blondel, Geoffrey Canright, Nicolas de Cordes, Sébastien Deletaille, Kenth Engø-Monsen, Manuel Garcia-Herranz, Jake Kendall, Cameron Kerry, Gautier Krings, Emmanuel Letouzé, Miguel Luengo-Oroz, Nuria Oliver, Luc Rocher, Alex Rutherford, Zbigniew Smoreda, Jessica Steele, Erik Wetter, Alex "Sandy" Pentland & Linus Bengtsson

*Scientific Data* 5, Article number: 180286 (2018) | [Download Citation](#) ↓

## LesEchos

DÉCRYPTAGE

## Nos données peuvent-elles servir l'intérêt général ?

PROSPECTIVE - De plus en plus de voix s'élèvent pour demander que les données collectées par les entreprises privées soient mises au service de la collectivité.



# “Open Algorithms” (OPAL): a revolutionary vision to make facts and measurements matter, and data, algorithms and AI ‘work’ for the majority and the SDGS

Tariq Khokhar @tkb · 2h  
Shoutout to the OPAL project - "bring the algorithm to the data" - more at: [opalproject.org](http://opalproject.org) #UNDataForum



**The Open Algorithm project:**  
Developing indicators, capacity and trust

To address the complex challenge of data access, Orange, MIT Media Lab, Data-Pop Alliance, Imperial College London and the World Economic Forum — supported by Agence Française de Développement and the World Bank — are developing a platform to unleash the power of “big data” held by private companies for public good in a privacy preserving, commercially sensible, stable, scalable and sustainable manner.

Open algorithms: A new paradigm for using private data for social good *By Thomas Roca, Emmanuel Letouzé | 18 July 2016*

Elisabeth MEDOU BA DANG,  
Porte-parole et directrice Afrique,  
Moyen-Orient, Orange  
Rabat, 2 juillet 2018

Aujourd’hui, le projet OPAL est développé avec l’Agence Française de Développement, Telefonika et d’autres partenaires. Des tests sont en cours au Sénégal et en Colombie. Nous espérons pouvoir mettre en œuvre plusieurs usages d’ici la fin de l’année. Cette plateforme constitue un levier important pour industrialiser l’usage du big data au service du développement.

COLOMBIE SÉNÉGAL 23 MAI 2018

PROJET OPAL : LE BIG DATA AU SERVICE DU DÉVELOPPEMENT



MONDE

## Mettre le Big Data privé au service du bien public

Le projet Open Algorithm vise à utiliser les données d’entreprises privées pour des actions de développement.

BENOÎT GEORGES - LES ECHOS | LE 06/12/2016



# OPAL: Setting Human AI data systems and standards for sharing and using private data safely at scale



1. Partner private companies (here a telecom operator) allow OPAL to access its servers through a secured platform. The **data never leave** the servers.

Global/Local/Crowd  
Open Algo check & certification



2. **Certified open algorithms** developed by developers are sent and run on the servers of partner private companies, behind their firewalls.

ALGO DEVELOPPERS

3. **Agovernance system** including a *Council for the Orientations of Development and Ethics (CODE)* ensures that the algorithms and use cases are ethically sound, context relevant, etc.; users benefit from **capacity building** activities

4. **Key indicators derived from private sector data** such as **population density, poverty levels, or mobility patterns, feed into use cases** in various public policy and economic domains. Data are safe, minimized, used (more) ethically.

Local  
C.O.D.E.



USERS





# OPAL is a unique case of a Public-Private-People Partnership piloted in Colombia and Senegal

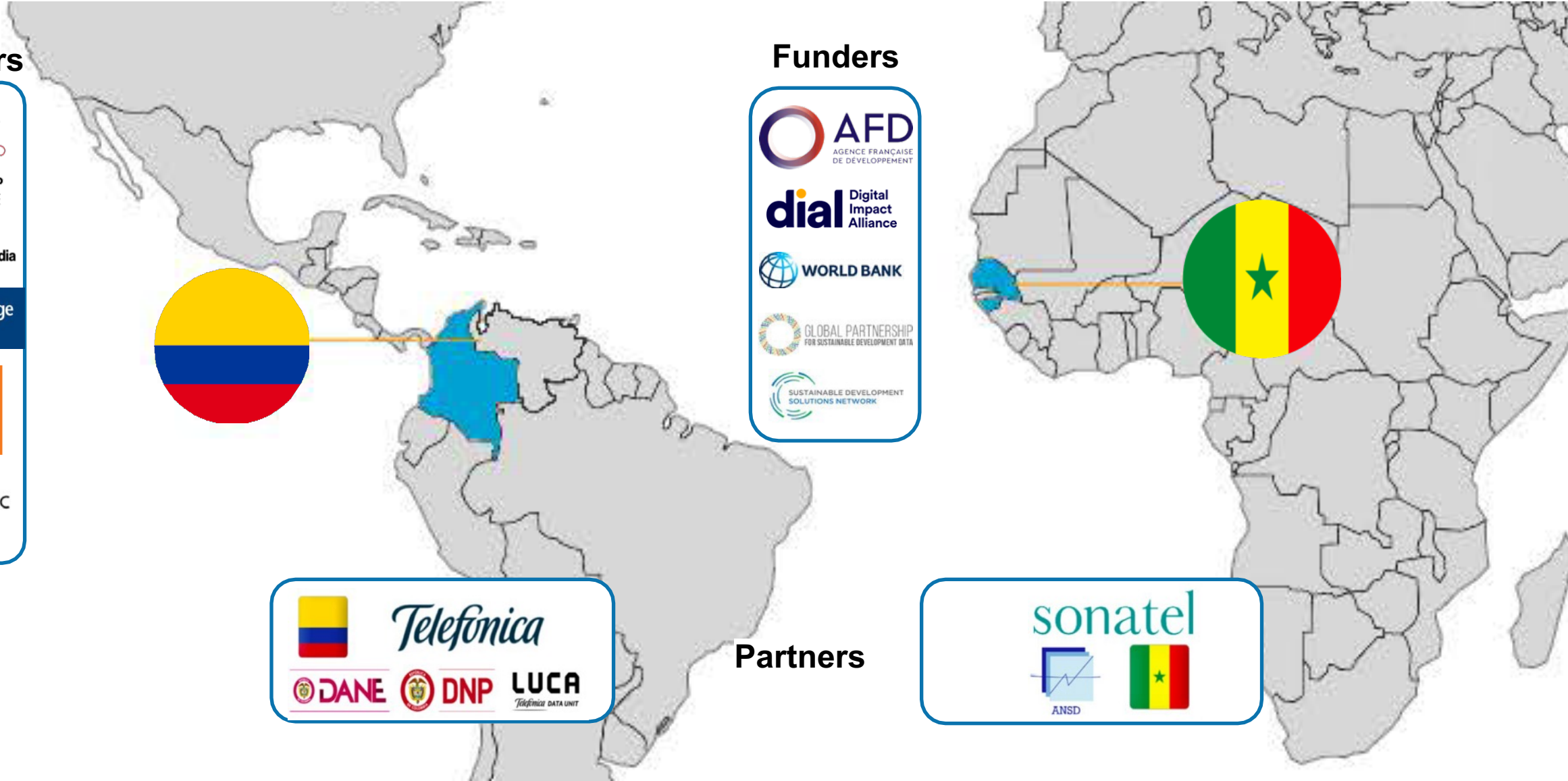
## Founders



## Funders



## Partners

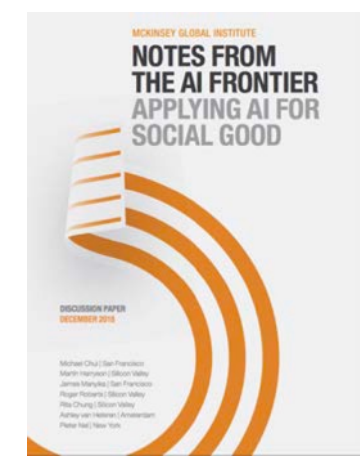
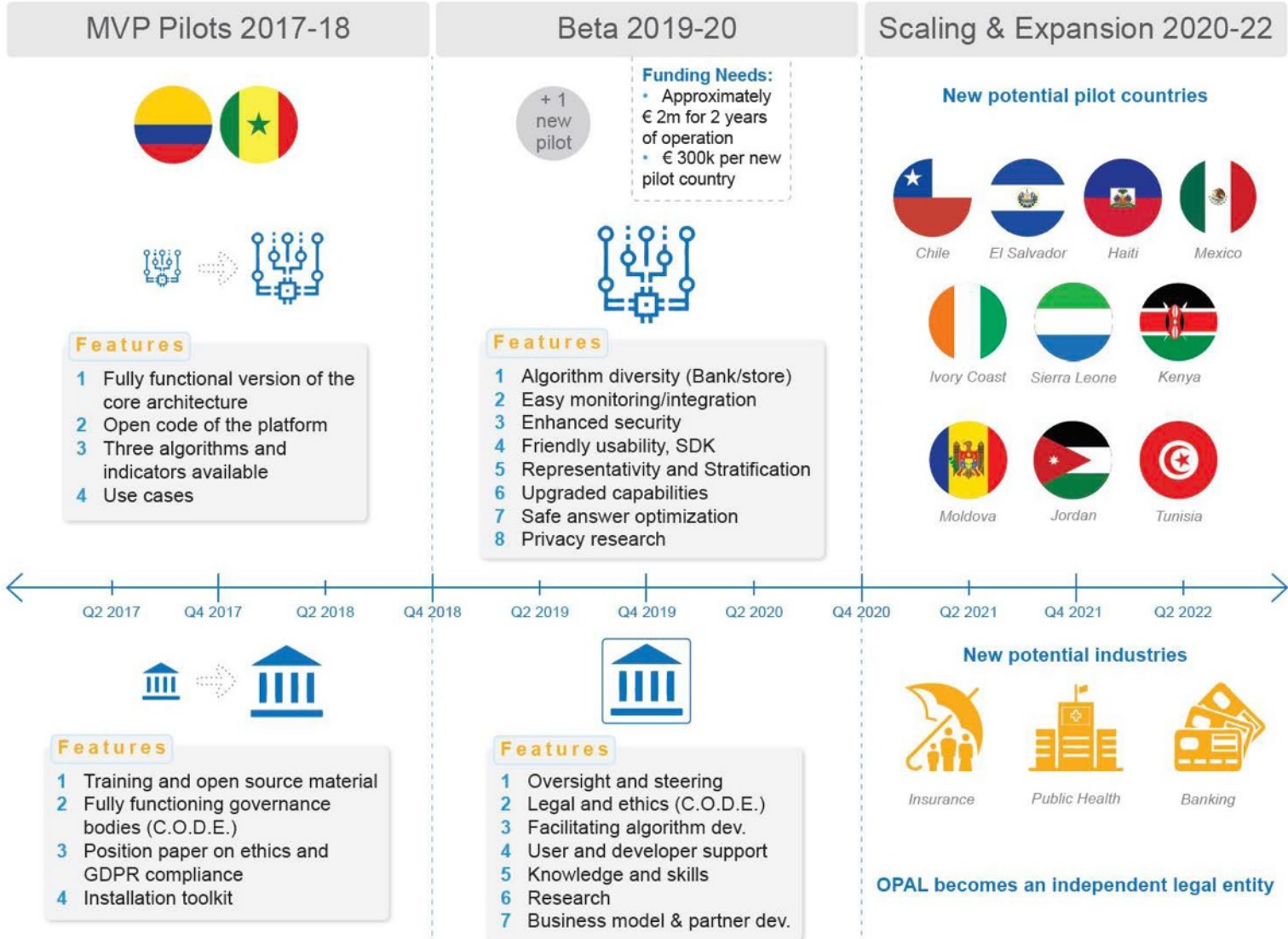




# OPAL ambitions to radically change how data are shared and used to improve the state of the world

TECHNOLOGY

GOVERNANCE



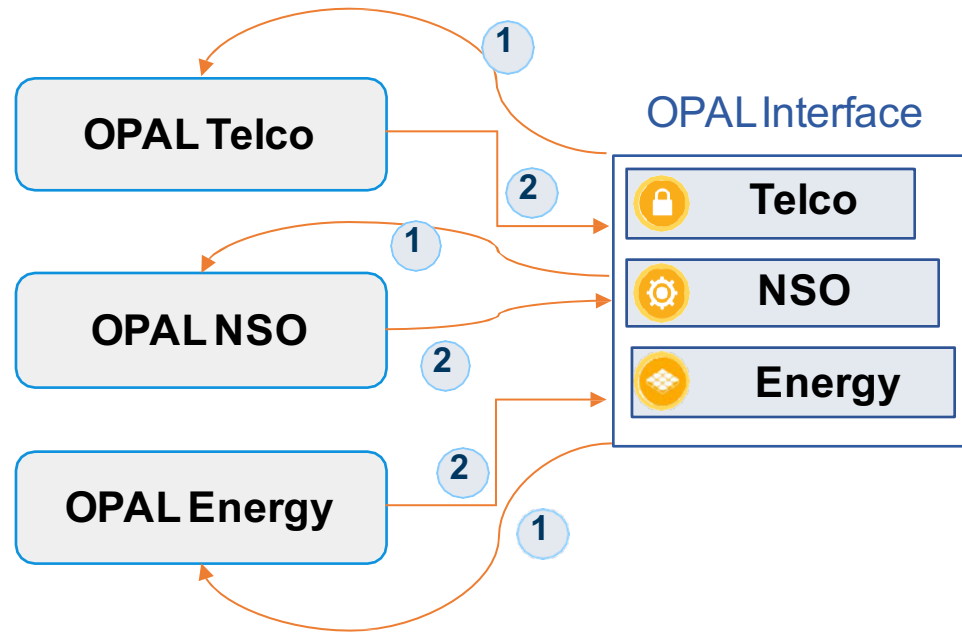
*“Other data sharing initiatives with private companies are also being worked on, including OPAL (...), to derive*

*aggregated insights from a company’s data without data leaving the company’s server. If proven successful, this could be a powerful tool in unlocking private data for social causes.”*

—McKinsey Global Institute *“Applying AI for Social Good”*, December 2018”

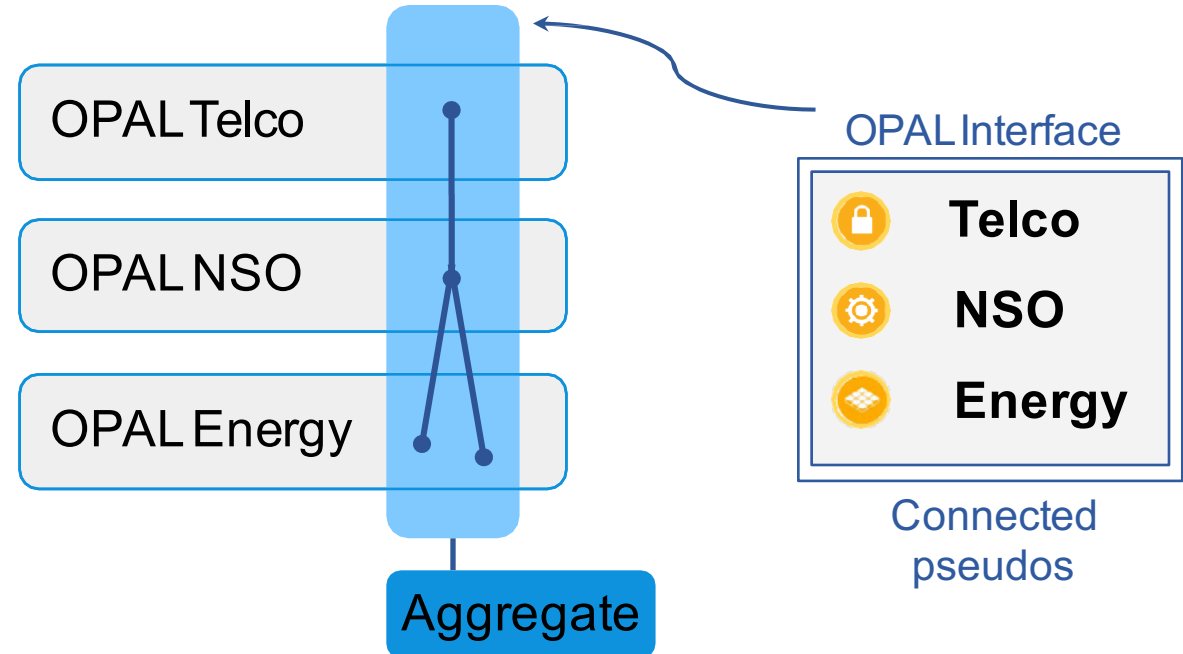
# OPAL will aim to query multiple data sources

**First step**  
Level 1: Additive



- 1 Call algo on separate OPALs
- 2 Get individually aggregated results (current MIT OPAL demo)

**OPAL's vision**  
Level 2: Integrative



- 1 Call level 2-algo that runs across OPALs
- 2 Aggregate combined results

# Use cases in Senegal



**Population Density**  
Calibration and use of population density algorithm (with **Knuper**)

**Health**  
Granularity and explanatory factors of contagious diseases



**Institut Pasteur**  
Analysis of Dengue *Essec-Accenture*

**Poverty**  
Analysis of monetary poverty



**Agriculture**  
Impact of road improvements on regional markets

**Agriculture**  
Analysis of travel time and distance to regional markets



**Urbanism**  
Analysis of mobility flows

**Economy**  
Event detection



**Migration and Stability**  
Migrations and tensions between communities

**Agriculture**  
Data Visualization and simulation tool

*Essec-Accenture*





# Use case #1 in Senegal

## Projections démographiques

Métadonnées de téléphone mobile pour  
les projections démographiques



Mars 2019

# How to build “data literacy”, connections, and “rational compassion” to make this work?

**DATA POP ALLIANCE**

**UNITED NATIONS SYSTEM STAFF COLLEGE**

**MOBILISER LE  
BIG DATA  
POUR LE  
DÉVELOPPEMENT DURABLE**

Atelier de formation professionnelle

**DAKAR**

**GUIDE DES PARTICIPANT-ES**

6 - 8 Mars, 2018  
Dakar, Sénégal



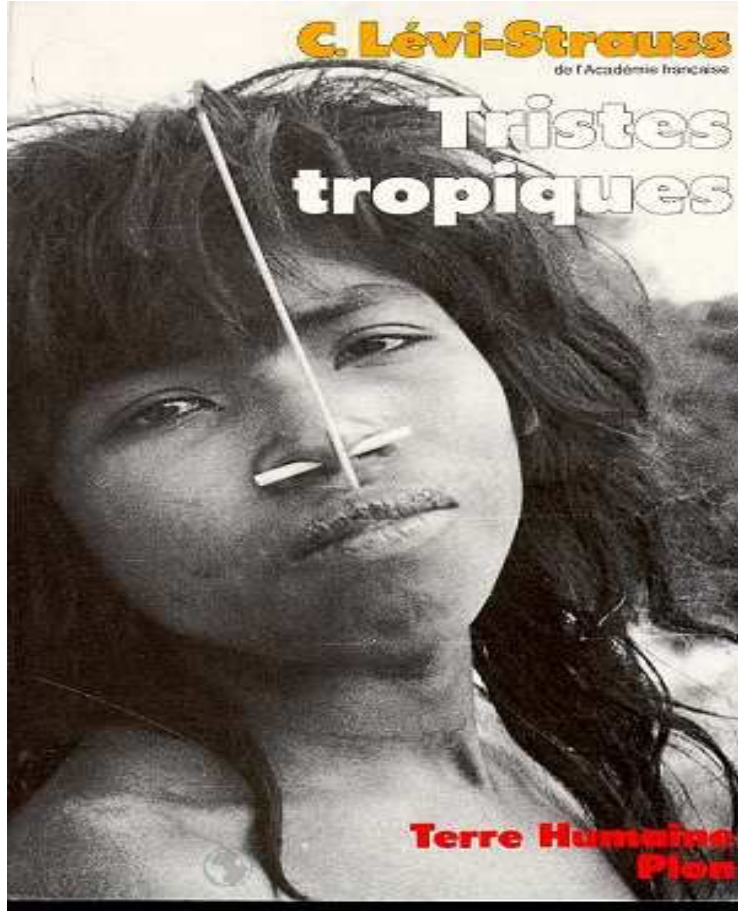
## Building Literacy for the Data Generation

*A unique opportunity exists to develop data literacy education for children born into a world shaped by big data.*

MasterCard Center for Inclusive Growth



# Think about the need for 'data literacy' and reconsider as literacy in the age of data



***"Writing is a strange thing. If my hypothesis is correct, the primary function of writing, as a means of communication, is to facilitate the enslavement of other human beings".***

***If writing was not sufficient to spur knowledge, it may have been necessary to reaffirm domination structures. (...) The fight against illiteracy goes on par with an increase in the control of the Power over citizens."***

Avoid this with data to **accelerate positive history**



# Think about the need for 'data literacy' and reconsider as literacy in the age of data

DATA-POP ALLIANCE  
WHITE PAPER SERIES

Beyond Data Literacy:  
Reinventing Community  
Engagement and Empowerment  
in the Age of Data

October 2015

*We define data literacy as the “the desire and ability to constructively engage in society through or about data”.*

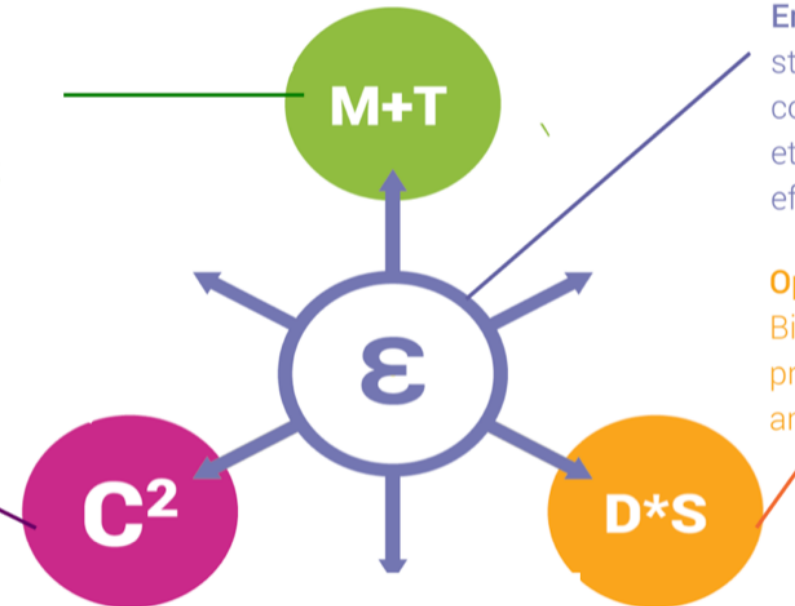
## Data-Pop Alliance's building blocks" of data literacy

Applying Big Data methods and tools to yield insights for specific development problems

Understanding key Big Data ideas in order to translate development problems into specific data objectives

Engaging key stakeholders and communities through ethical practices and effective story-telling

Operationalizing Big Data as inclusive projects, partnerships and policies



## Building Literacy for the Data Generation

December 18, 2015



MasterCard Center  
for Inclusive Growth

*A unique opportunity exists to develop data literacy education for children born into a world shaped by big data.*

## Trainings Timeline

**CAMBRIDGE, MA**

June 2016

**BOGOTÁ, COLOMBIA**

December 2016

[Learn more](#)

**SANTIAGO DE CHILE, CHILE**

March 2017

**NAIROBI, KENYA**

June 2017

[Learn more](#)

**SÃO PAULO, BRASIL**

September 2017

[Learn more](#)

**MEXICO CITY, MEXICO**

October 2017

[Learn more](#)

**DAKAR, SENEGAL**

March 2018

**CAMBRIDGE, MA (MIT)**

October 2018

[Learn more](#)

**TUNIS, TUNISIA**

April 2019

[Learn more](#)

**BANGKOK, THAILAND**

March 2018

**SANTO DOMINGO, DOMINICAN REPUBLIC**

April 2019

[Learn more](#)

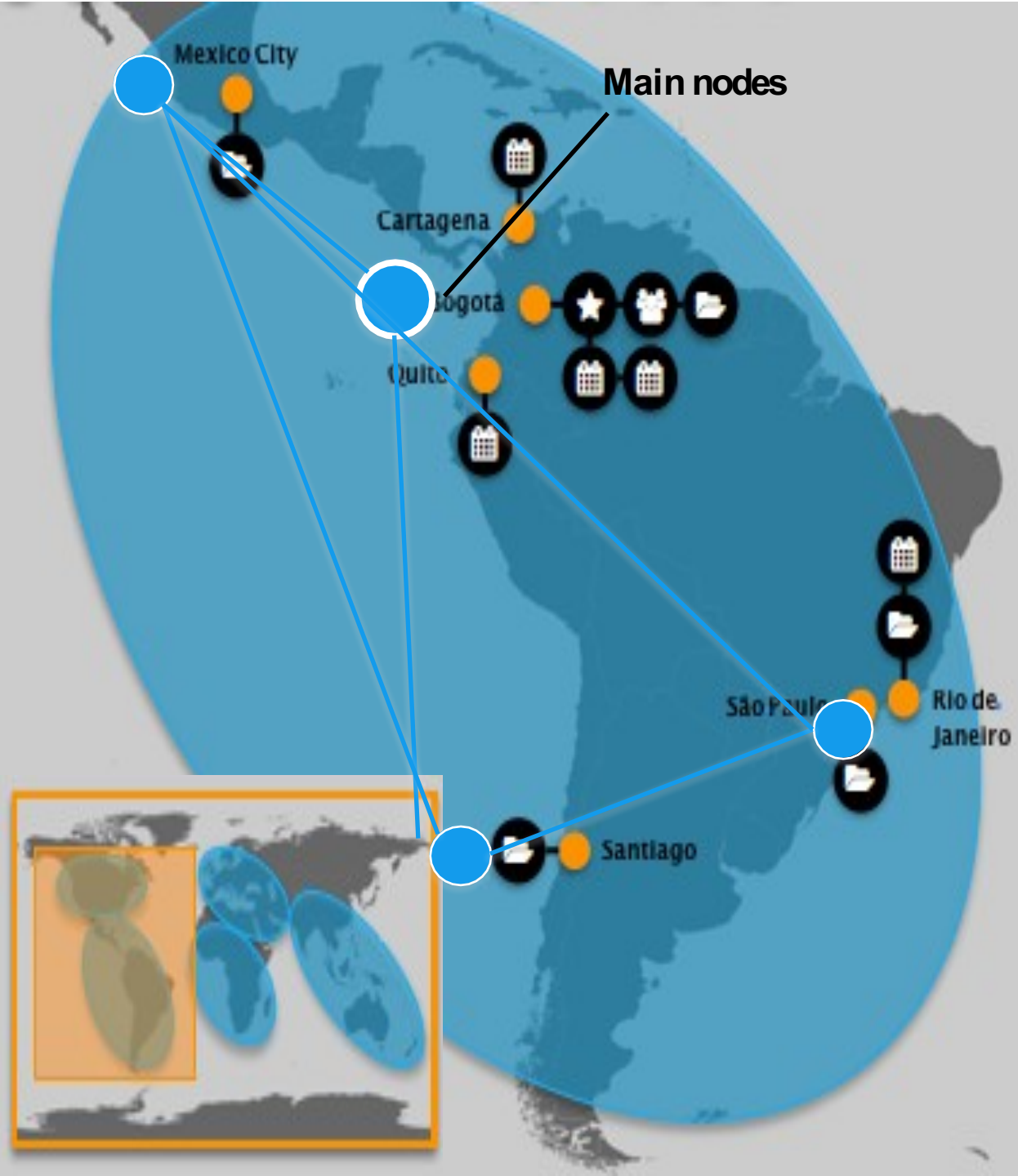
**BOGOTÁ, COLOMBIA**

May 2019

# Data-Pop work and model

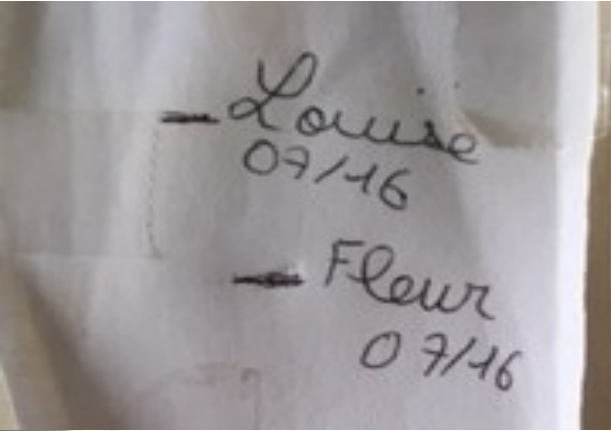
(here example of LAC):

*Locally co-design and deploy regional research, training, and strategy programs and partnerships to leverage Big Data and AI for sustainable human development (including outside of LAC in Botswana, Togo, Turkey, Moldova, Turkey, Tunisia...)*

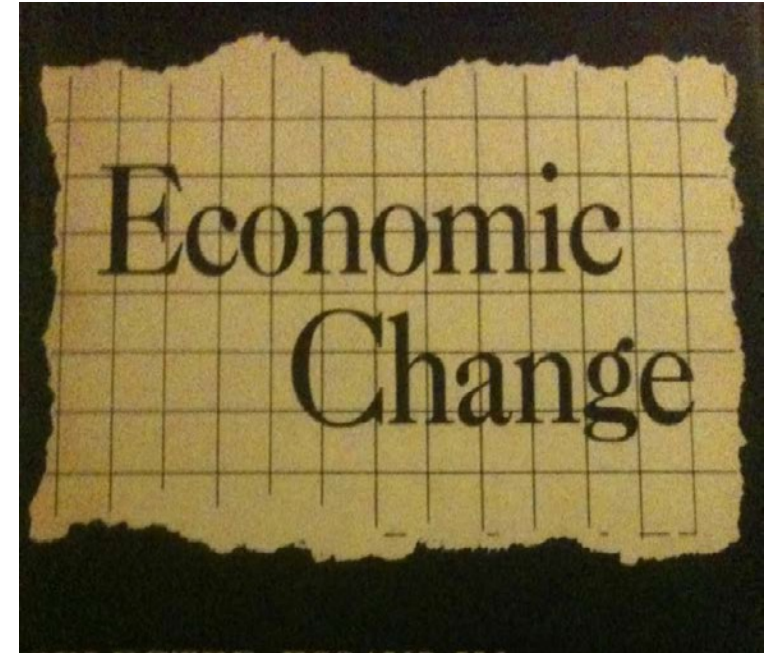




# Why do we measure things? Does it matter? Why or why not? How can it matter more?

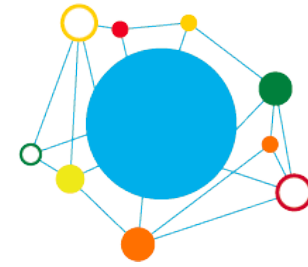


Measurement only matters when people care. Measurement is meant to signal and strengthen care about what gets measures; to turn codes into norms. Typically, in simple systems measurement becomes unnecessary once people care about what is measured. In complex system like human societies, the language, appetite for, skills and culture of measurement used and promoted wisely and with “rational compassion”, remains a very powerful tool to instill change.





MIT  
Connection  
Science



DATA-POP  
ALLIANCE

# Thank you

[eletouze@datapopalliance.org](mailto:eletouze@datapopalliance.org)

[@opalproject.org](https://opalproject.org)

[@mit.edu](https://mit.edu)

IX NIC.br Annual  
**Workshop  
on Survey  
Methodology**