

Emmanuel Letouzé, PhD

Data, Development, and Democracy?
Leveraging Big and Open Data
to Build Safer and Fairer Societies

University Pompeu Fabra

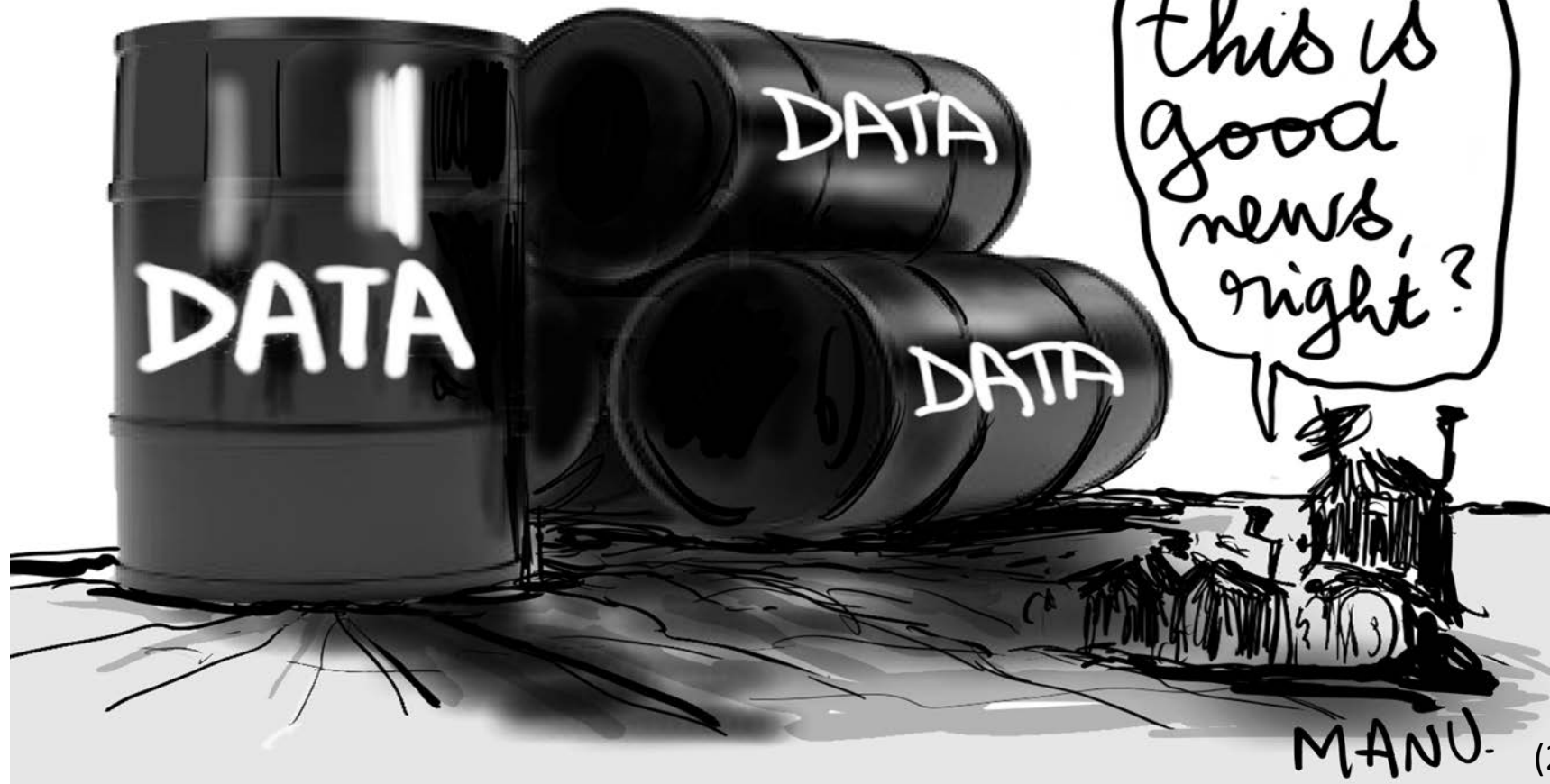
Barcelona

September 13, 2017

Thanks

John Palmer, Gosta Erding-Andersen,
Aïda Solé Auró, Maria Jose Gonzales
Lopez....

Data is the new oil



MANU. (2011)



Can Big Data From Cellphones Help Prevent Conflict?

November 8, 2012 • by Emmanuel Letouzé



Data from social media and Ushahidi-style crowdsourcing platforms have emerged as possible ways to leverage cellphones to prevent conflict. But in the world of Big Data, the amount of information generated from these is too small to use in advanced data-mining techniques and “machine-learning” techniques (where algorithms adjust themselves based on the data they receive).

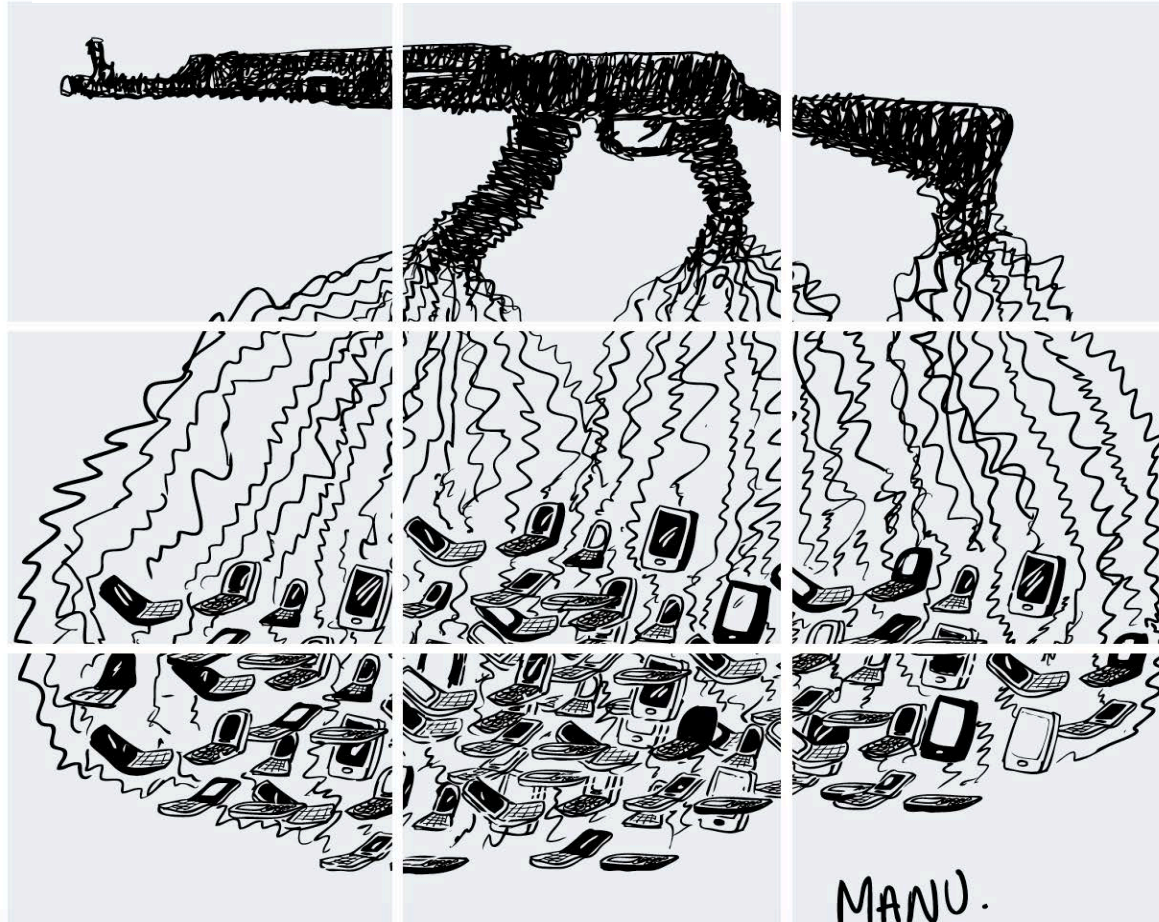
But there is another way cellphones could be leveraged in conflict settings: through the various types of data passively generated every time a device is used. “Phones can know,” said Professor Alex “Sandy”

Pentland, head of the Human Dynamics Laboratory and a prominent computational social scientist at MIT, in a [Wall Street Journal article](#). He says **data trails left behind** by cellphone and credit card users—“digital breadcrumbs”—reflect actual behavior and can tell objective life stories, as opposed to what is found in social media data, where intents or feelings are obscured because they are “edited according to the standards of the day.”

New Technology and the Prevention of Violence and Conflict

EDITED BY FRANCESCO MANCINI

APRIL 2013



*Empowered lives.
Resilient nations.*



USAID
FROM THE AMERICAN PEOPLE

IPI

INTERNATIONAL
PEACE
INSTITUTE

CONTENTS

Executive Summary	iii
Introduction.....	1
<i>Francesco Mancini</i>	
Big Data for Conflict Prevention: New Oil and Old Fires	4
<i>Emmanuel Letouzé, Patrick Meier, and Patrick Vinck</i>	
Using Information and Communication Technologies for Violence Prevention in Latin America.....	28
<i>Robert Muggah and Gustavo Diniz</i>	
Early Warning and the Role of New Technologies in Kenya	42
<i>Godfrey M. Musila</i>	
Conflict Cure or Curse? Information and Communication Technologies in Kyrgyzstan.....	56
<i>Anna Matveeva</i>	
New Technologies and Conflict Prevention in Sudan and South Sudan.....	71
<i>Helena Puig Larrauri</i>	
Conclusion: New Technology in Conflict Prevention	87
<i>Francesco Mancini and Marie O'Reilly</i>	

Big Data for Conflict Prevention: New Oil and Old Fires¹

Emmanuel Letouzé, Patrick Meier, and Patrick Vinck²

The ability to manipulate big data, visualize dynamics, and recognize patterns and signatures for conflict creates new opportunities for humanitarian and development assistance in the most complex and dangerous environments.

David Kilcullen and Alexa Courtney³

The theory of technology as amplifier explains how the same technology can appear to have both positive and negative impacts, because technology is merely a magnifier of underlying human and institutional intent and capacity, which can themselves be positive or negative.

Kentaro Toyama⁴

The hope that technology will reduce the violence of war is a venerable one... Richard Gatling hoped his new fast-firing gun would serve to reduce the bloodshed of war, while Alfred Nobel believed the explosives he invented would make war unthinkable.

Peter W. Singer⁵

**Does/Will (Big) Data Entrench Elites,
or Empower People?**

Does/Will (Big) Data Entrench Elites,
or Empower People?
To What End?

Does/Will (Big) Data Entrench Elites,
or Empower People?
To What End?
How?

**(How) Can (Big) Data (Help) Build
Fairer and Safer Future Societies?**

CYBERSECURITY

Saving Big Data from Itself

A three-step plan for using data right in an age of government overreach

By Alex "Sandy" Pentland

For the first few decades of its existence, the National Security Agency was a quiet department with one primary job: keeping an eye on the Soviet Union. Its enemy was well defined and monolithic. Its principal tools were phone taps, spy planes and hidden microphones.

After the attacks of September 11, all of that changed. The NSA's chief enemy became a diffuse network of individual terrorists. Anyone in the world could be a legiti-



CULTURE | SOCIOLOGY

To Rescue Democracy, Go Outside

Real spaces, not digital ones, will fix our politics.

BY ALEX PENTLAND
ILLUSTRATION BY JAMES YANG
OCTOBER 13, 2016

SCIENTIFIC
AMERICAN™

Saving Big Data from Big Mouths

Those who would condemn big data ought to try making something

By Cesar A. Hidalgo | April 29, 2014

SA Forum is an invited essay from experts on topical issues in science and technology.

It has become fashionable to bad-mouth big data. In recent weeks the *New York Times*, *Financial Times*, *Wired* and other outlets have all run pieces bashing this new technological movement. To be fair, many of the critiques have a point: There has been a lot of hype about big data and it is important not to inflate our expectations about what it can do.



There has been a lot of hype about big data and it is important not to inflate our expectations about what it can do.

Credit: DARPA via Wikimedia Commons

“A ‘true’ big data revolution should be one where data can be leveraged to change power structures and decision-making processes, not just create insights.”

Emmanuel Letouzé

Applications and Implications of Big Data for Demo-Economic Analysis:
The Case of Call-Detail Records

by

Emmanuel Francis Letouzé

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy
in
Demography
in the
Graduate Division
of the
University of California, Berkeley

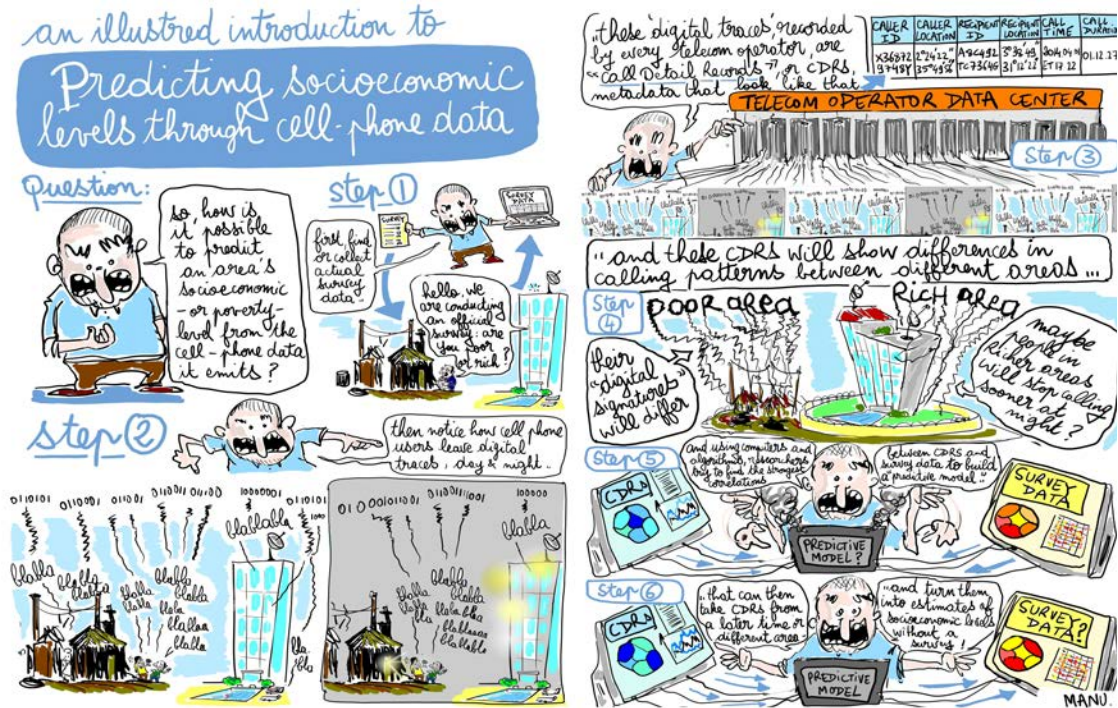
Committee in charge:

Professor Ronald Lee, Chair
Professor Jennifer Johnson-Hanks
Professor Edward Miguel

Summer 2016

CHAPTER 1

Figure 1.6: An illustrated introduction to predicting socio-economic levels through cell-phone data



A Decade of Big Data Rush

"We are at the beginning of what I call The Industrial Revolution of Data."
Joe Hellerstein,
November 19, 2008

Big data: The next frontier for innovation, competition, and productivity

Big Data for Development: Challenges & Opportunities
May 2012

BIG DATA
A REVOLUTION THAT WILL TRANSFORM HOW WE LIVE, WORK, AND THINK

SOCIAL PHYSICS
HOW SOCIAL NETWORKS CAN MAKE US SMARTER
ALEX PENTLAND

A WORLD THAT COUNTS



GLOBAL PARTNERSHIP FOR SUSTAINABLE DEVELOPMENT DATA

WIRED MAGAZINE: 16.07

The End of Theory: The Data Deluge Makes the Scientific Method Obsolete

WORLD ECONOMIC FORUM

Big Data, Big Impact: New Possibilities for International Development

The Economist

Unique in the Crowd: The privacy bounds of human mobility

Yves-Alexandre de Montjoye^{1,2}, César A. Hidalgo^{1,3,4}, Michel Verleyen⁵ & Vincent D. Blondel^{1,3}

Data and development
Off the map

WEAPONS OF MATH DESTRUCTION
CATHY O'NEIL

50

SOCIAL SCIENCE

Computational Social Science

David Lazear,¹ Alex Pentland,² Lada Adamic,³ Sinan Aral,^{2,4} Albert-László Barabási,⁵ Devon Brewer,⁶ Nicholas Christakis,¹ Noshir Contractor,⁷ James Fowler,⁸ Myron Gutmann,² Tony Jebara,⁹ Gary King,¹ Michael Macy,¹⁰ Deb Roy,² Marshall Van Alstyne^{2,11}

Africa's statistical tragedy

Cell-Phone Data Might Help Predict Ebola's Spread

25

Statistics 2.0
The next level

By Enrico Giovannini

Is data a danger to the developing world?

The Tyranny of Data?
The Bright and Dark Sides of Data-Driven Decision-Making for Social Good

Bruno Lepri, Jacopo Staiano, David Sangokoya, Emmamel Letouzé and Nuria Oliver

02 Nov 2015
Kate Crawford
Principal Researcher, Microsoft Corp

Interest / returns for « Big Data » on Google (2007-2017)





60th Anniversary of the National
Department of Statistics of Colombia

**Official Statistics
in the Big Data Era**

Emmanuel Letouzé

PhD Candidate, UC Berkeley
Fellow, Harvard Humanitarian Initiative

eletouze@Berkeley.edu

Biblioteca Luis Angel Arango
Bogotá October 28th, 2013

MANU.

(2011)

i just heard the founding fathers rolling over in their graves!



Technology

Data for Development

Associated Channels

Technology

Governance

Stratification

April 23, 2015 | by *Joy Neumeyer*



SHARE

UC Berkeley demographer (and cartoonist) Emmanuel Letouzé seeks to reduce the gap in statistical data available about populations in developing nations.



Technology

Data for Development

Associated Channels

Technology

Governance

Stratification

April 23, 2015 | by *Joy Neumeyer*

SHARE

Poverty and underdevelopment often go hand-in-hand with a lack of data, as regions lacking infrastructure or mired in violent conflict are often unable to conduct surveys and compile statistics about populations. In Afghanistan or the Democratic Republic of Congo, for example, a [census](#) has not been conducted in decades. And while in the U.S., fitness trackers and Google Maps track people's most incremental movements, in Africa, many births go unrecorded, and online maps show only major cities and streets. As a result, vital information about developing nations—such as poverty, health, and unemployment levels—often can only be estimated. Even initiatives such as the [UN's Millenium Development Goals](#), which focuses on child poverty, health, and education, have struggled to accurately measure whether targets have been met.



1 NO POVERTY

A white silhouette icon of a family consisting of two adults and two children.

2 ZERO HUNGER

A white silhouette icon of a bowl with three wavy lines above it representing steam or smoke.

3 GOOD HEALTH AND WELL-BEING

A white silhouette icon of a heartbeat line with a heart symbol at the end.

4 QUALITY EDUCATION

A white silhouette icon of an open book and a pencil.

5 GENDER EQUALITY

A white silhouette icon of the gender equality symbol, which is a circle with an equals sign inside, and a cross and a triangle pointing outwards from the top right.

6 CLEAN WATER AND SANITATION

A white silhouette icon of a water tap with a single drop of water falling from it.

7 AFFORDABLE AND CLEAN ENERGY

A white silhouette icon of a sun with a power button symbol in the center.

8 DECENT WORK AND ECONOMIC GROWTH

A white silhouette icon of a bar chart with an upward-pointing arrow.

9 INDUSTRY, INNOVATION AND INFRASTRUCTURE

A white silhouette icon of three stacked cubes.

10 REDUCED INEQUALITIES

A white silhouette icon of an equals sign inside a circle.

11 SUSTAINABLE CITIES AND COMMUNITIES

A white silhouette icon of several buildings of different heights.

12 RESPONSIBLE CONSUMPTION AND PRODUCTION

A white silhouette icon of an infinity symbol.

13 CLIMATE ACTION

A white silhouette icon of an eye with a globe of the Earth as the pupil.

14 LIFE BELOW WATER

A white silhouette icon of three wavy lines representing water and a fish below them.

15 LIFE ON LAND

A white silhouette icon of a tree with two birds flying above it.

16 PEACE AND JUSTICE STRONG INSTITUTIONS

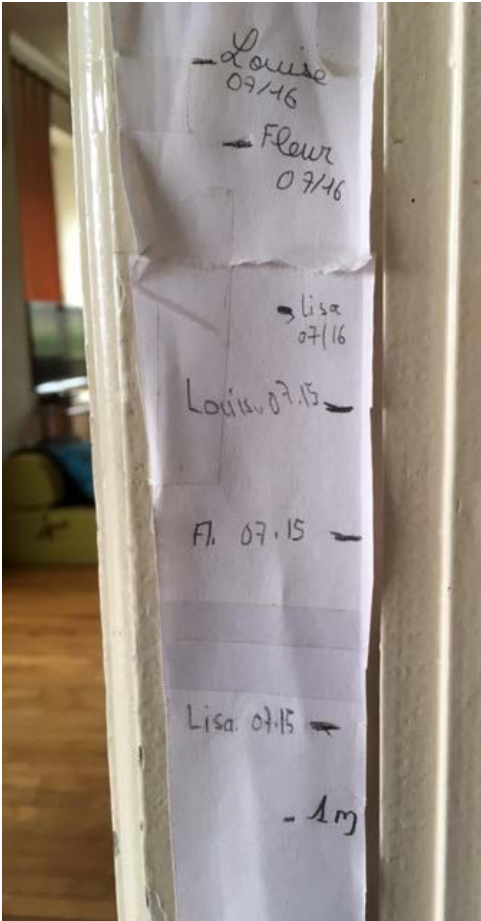
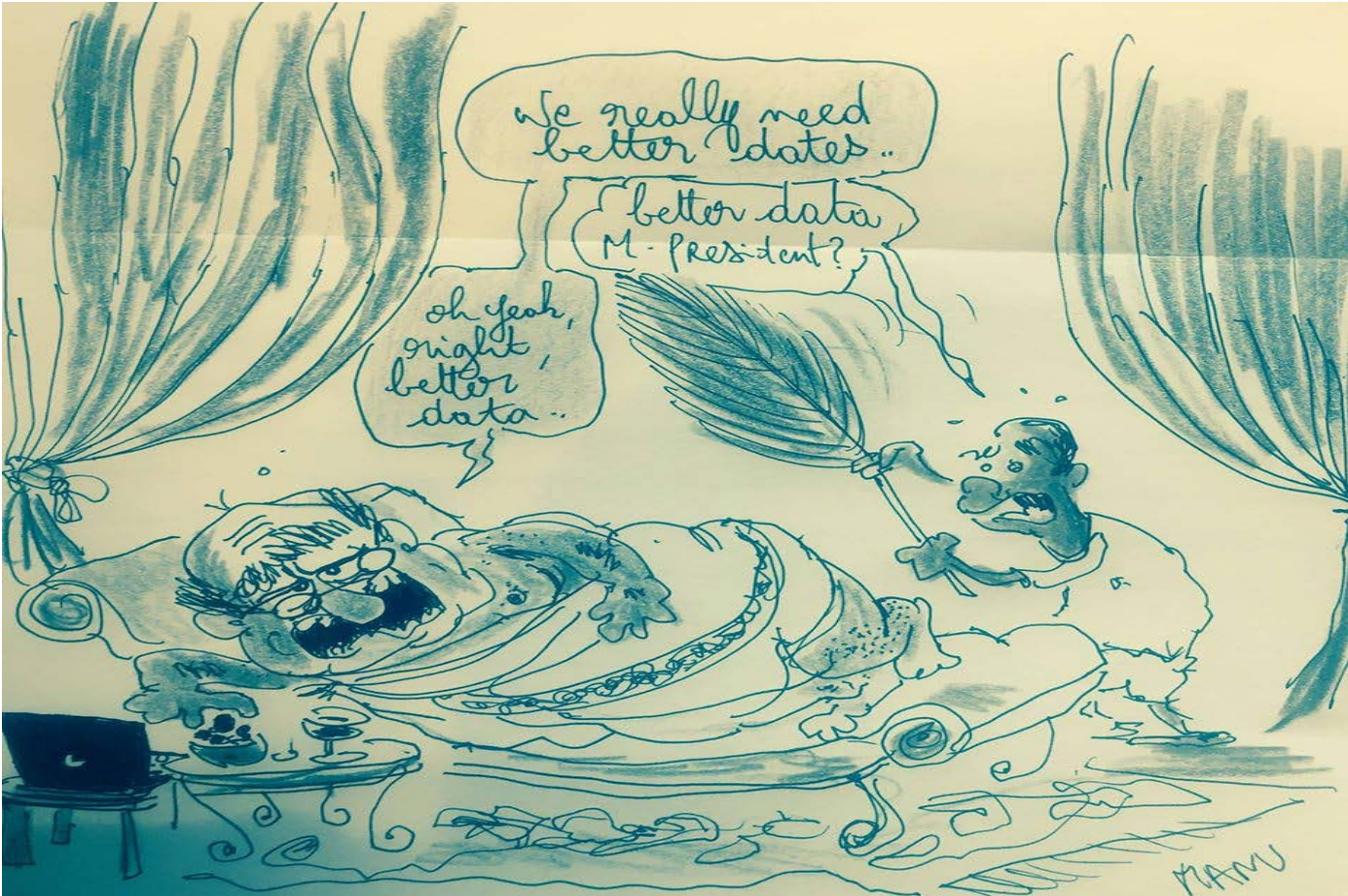
A white silhouette icon of a dove holding an olive branch and a gavel.

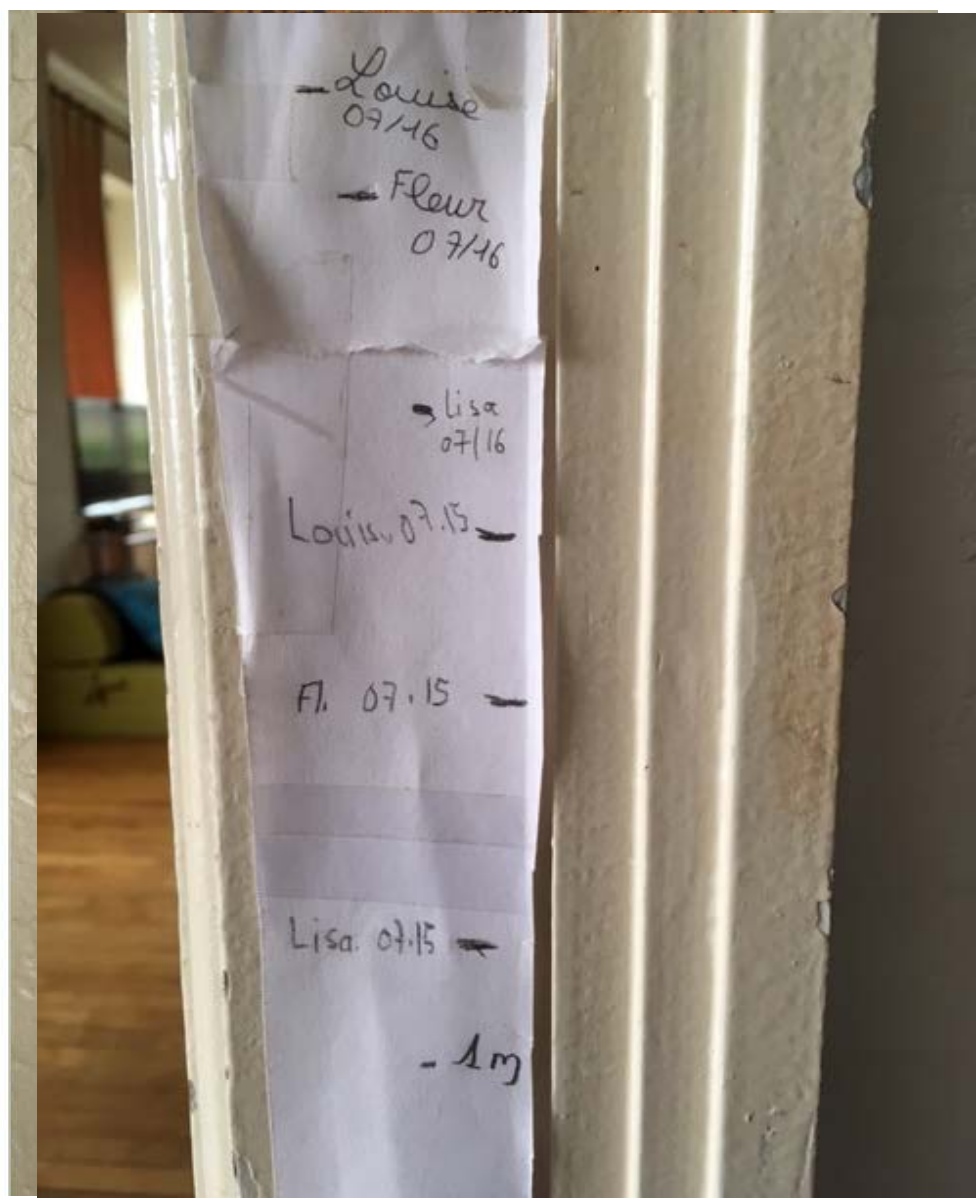
17 PARTNERSHIPS FOR THE GOALS

A white silhouette icon of five interlocking circles.

We Have Goals. Do We Have a Plan?

Poverty == f (Poverty Data?)







Trust
me

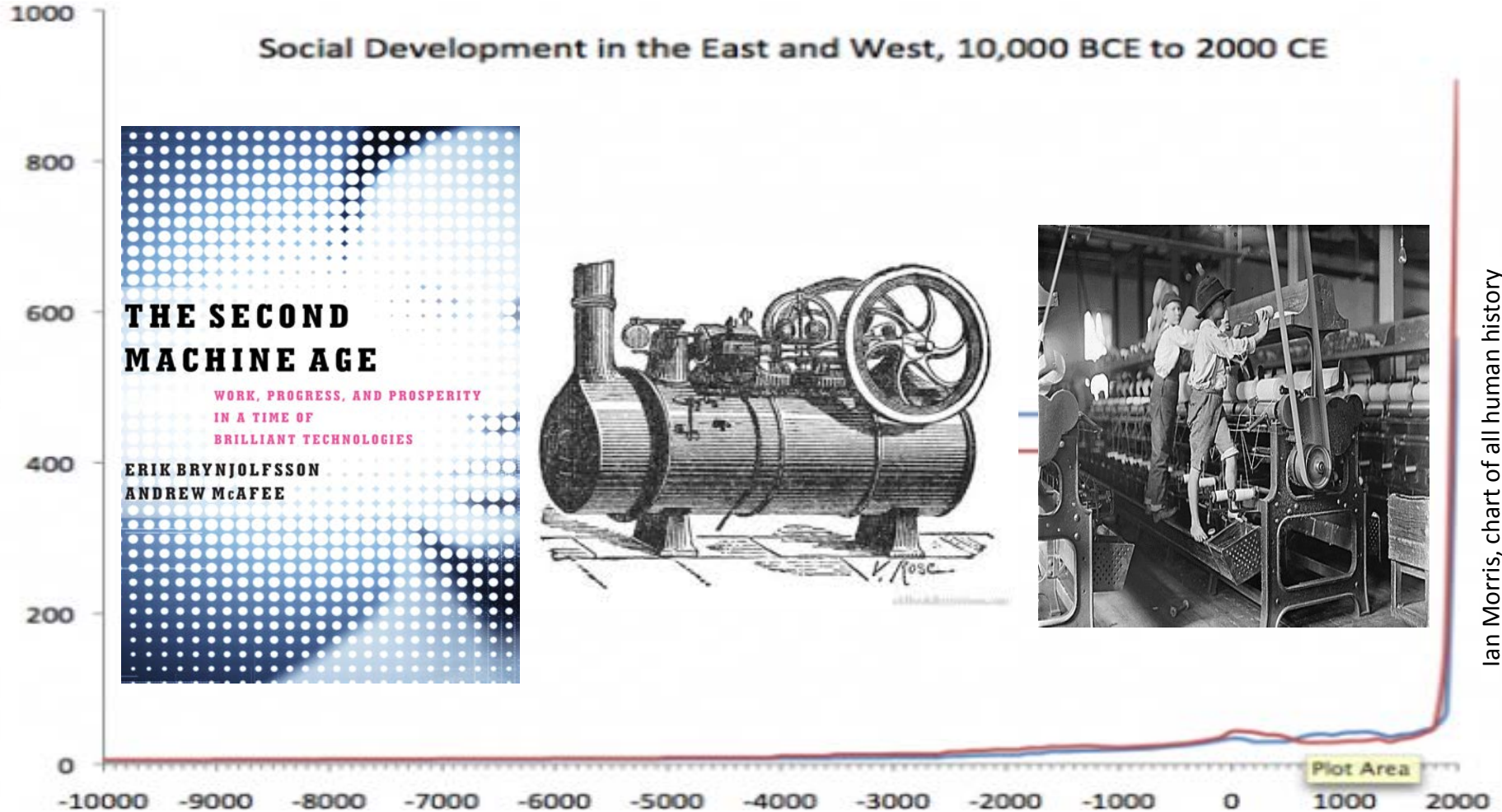


i'm
a data scientist

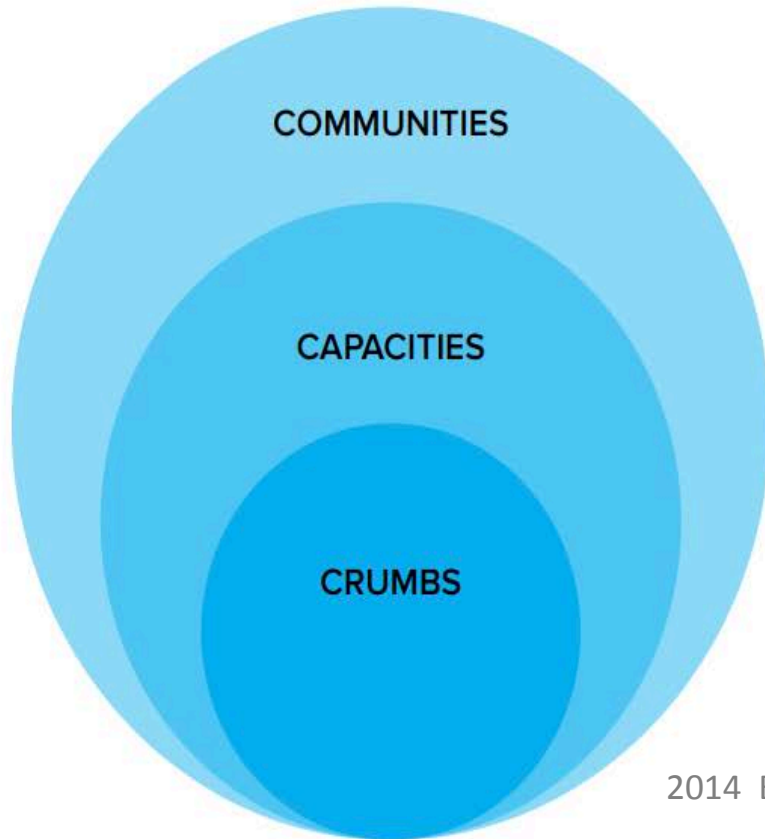


Bogo

Learning from the Lessons of History



From the 3Vs of big data as data to the 3 Cs of Big Data *as an ecosystem*



Originally framed as the “3 V’s” (volume, velocity and variety) in the early 2000s, Big Data has emerged as an ecosystem of “3 C’s”: digital “crumbs” (digital translations of human actions and interactions captured by digital devices); powerful capacities to collect, aggregate and analyze data; and communities involved in generating, governing and using data, including data generators, end users, policy-makers, experts, privacy advocates and civic hacker communities.



MIT Media Lab

@medialab

Following



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



Solving the data conundrum: How to leverage tech and 'big data' for impact

From combing phone subscription records to estimate population density and poverty levels, to analyzing tweets to predict a pending food crisis, emerging tec...

[devex.com](#)

RETWEETS

16

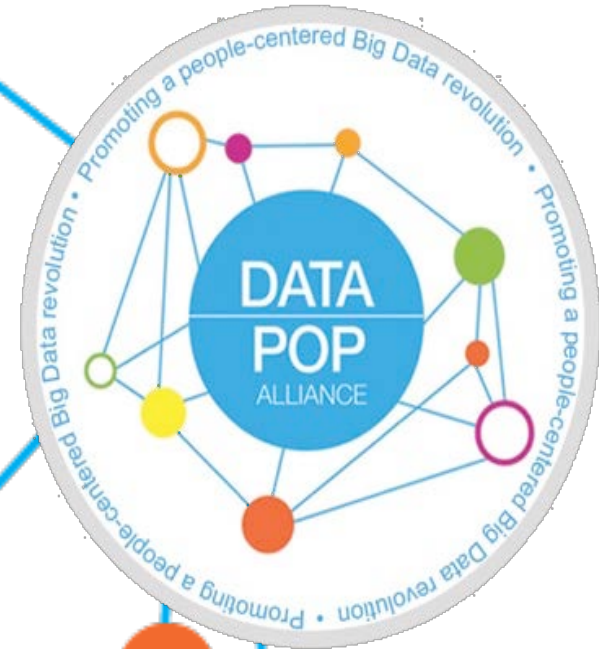
LIKES

20



12:08 PM - 18 Jul 2016

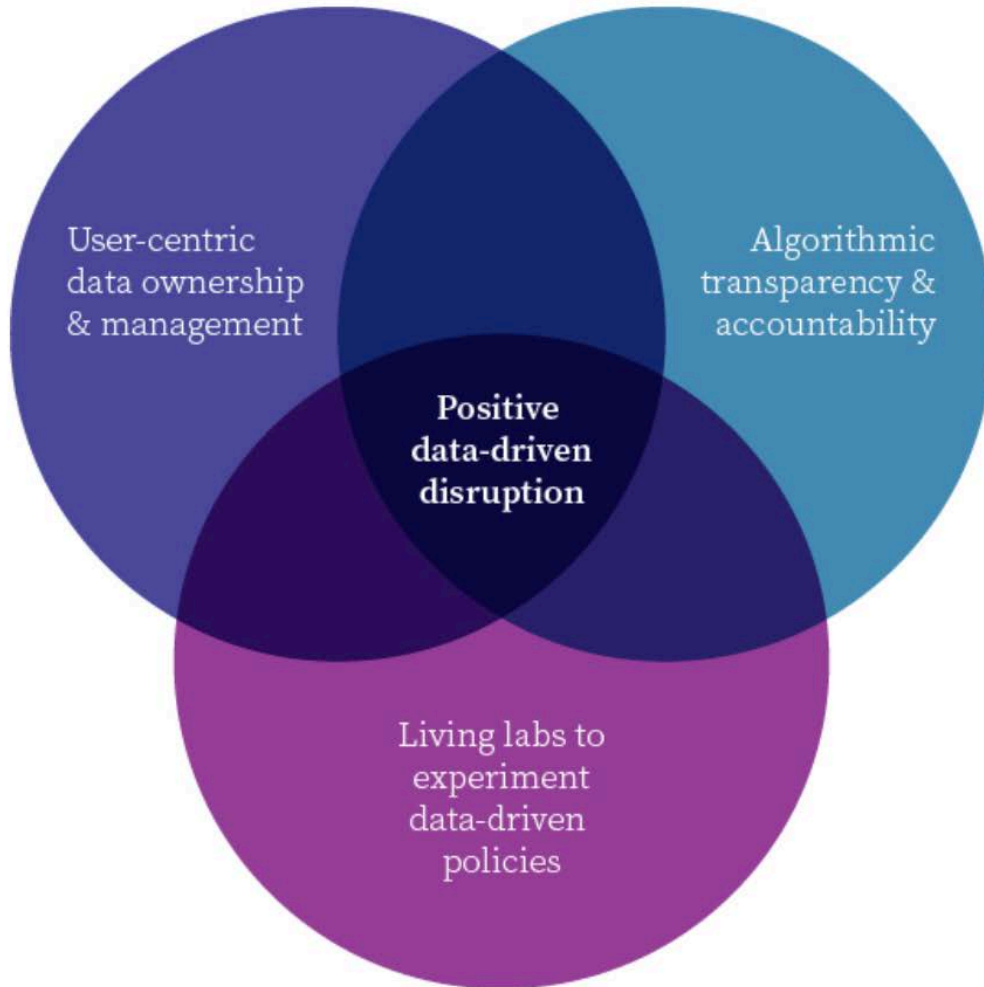
Data-Pop Alliance is a global coalition on Big Data and development created in 2014 by the Harvard Humanitarian Initiative, MIT Media Lab, and Overseas Development Institute, joined by Flowminder, bringing together researchers, experts, practitioners, and activists to promote a people-centered Big Data revolution



THE
ROCKEFELLER
FOUNDATION

THE WILLIAM AND FLORA
HEWLETT
FOUNDATION

Requirements/Pathways?

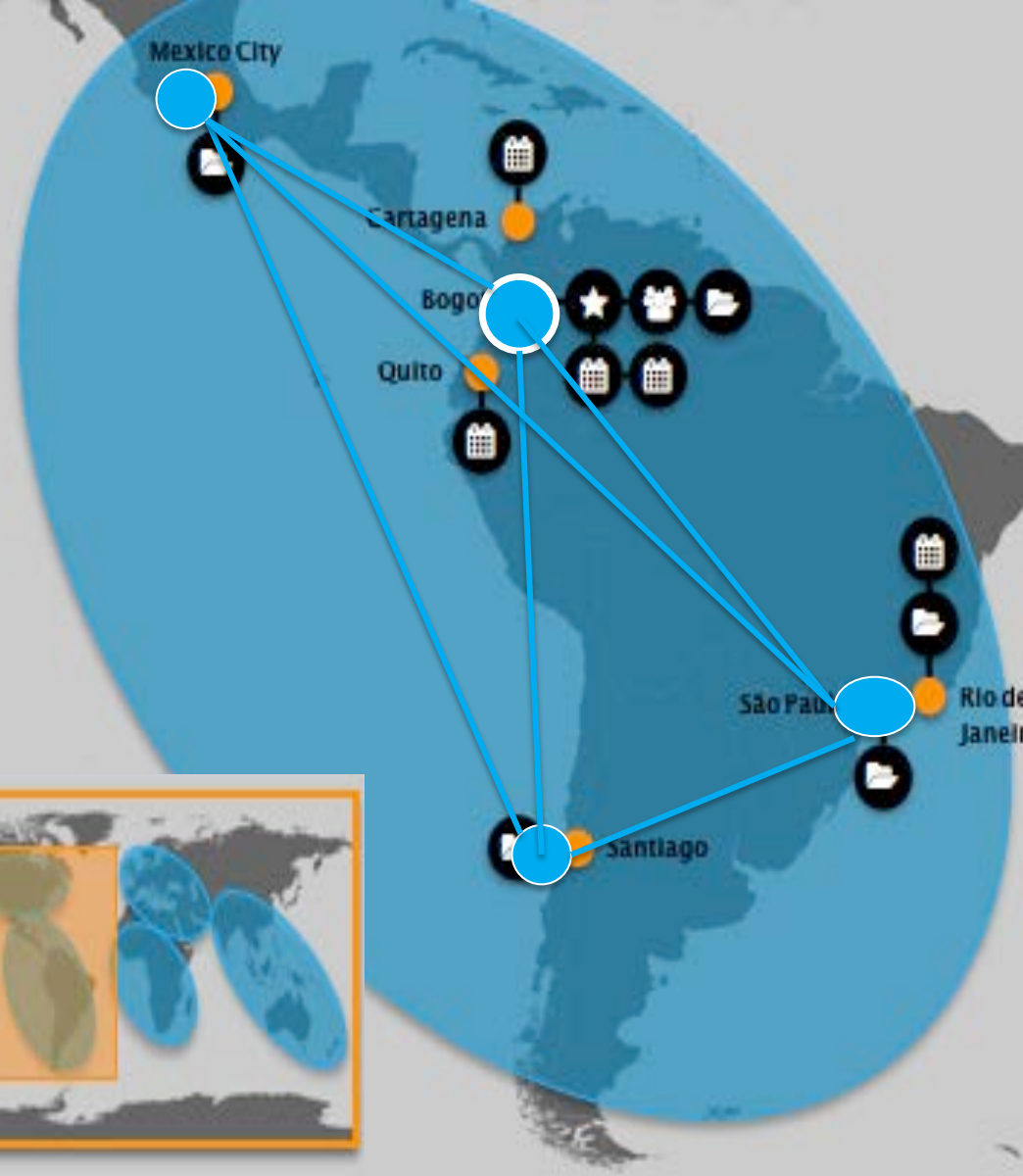


The Tyranny of Data?
The Bright and Dark Sides of
Data-Driven Decision-Making for
Social Good

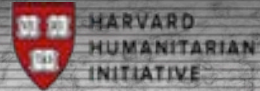
Bruno Lepri, Jacopo Staiano, David Sangokoya, Emmanuel Letouzé and
Nuria Oliver

*“It is only when we honor
theserequirements that we
will be able to move from the
feared tyranny of data and
algorithms to a data-enabled
model of democratic
governance running against
tyrants and autocrats, and for
the people.” ...?*

A Focus on LAC



Project #1: Bogotá *Laboratorio Urbano*



ALCALDÍA MAYOR DE BOGOTÁ D.C.

SECRETARÍA DE DESARROLLO ECONÓMICO

BOGOTÁ MEJOR PARA TODOS



Project #1: Bogotá *Laboratorio Urbano*



EXPLORAR

CARTOGRAFÍA

API

FORO

Bienvenido al **Laboratorio Urbano - Bogotá** portal de datos

/ buscar datos

Economía

Población

Comercio

Ambiente

Cultura, deport

Salud

Educación

Tecnología

Proyecciones

Culture, Sports

Economy

Encuestas

Population

Prices, Costs

Prices, Costs

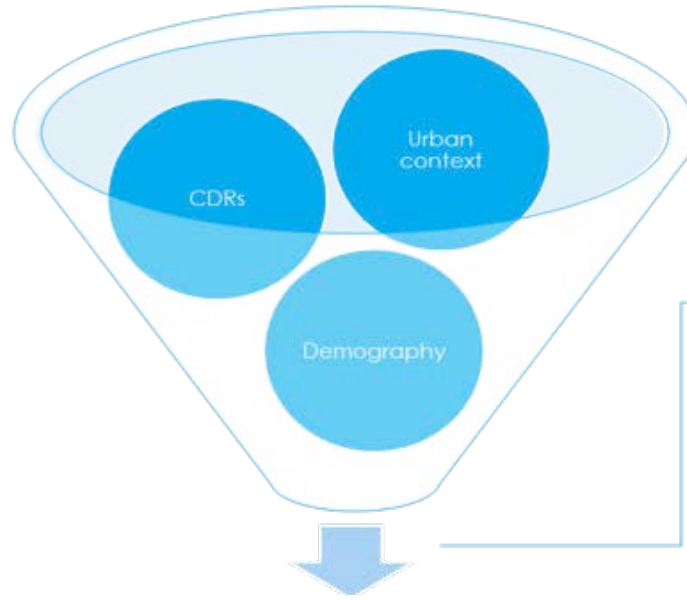
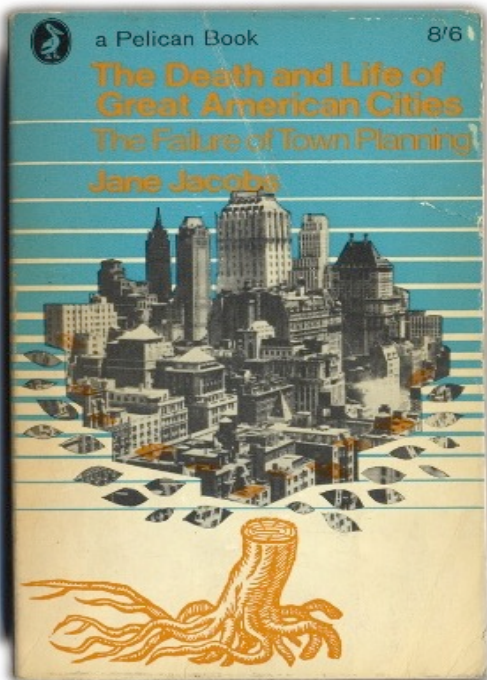
Surveys

Project #2: Urban Safety and Vitality

Inside Out: to understand crime mechanisms look at urban fabric first

Marco De Nadai^{*1,2}, Emmanuel Letouzé^{†4}, Marta C. González^{‡3} and Bruno Lepri^{§2}

¹University of Trento. ²FBK. ³MIT. ⁴Data-Pop Alliance



“Preliminary findings indicate that Jacobs' diversity conditions are a better predictor of homicides and robberies than socio-economic conditions.”

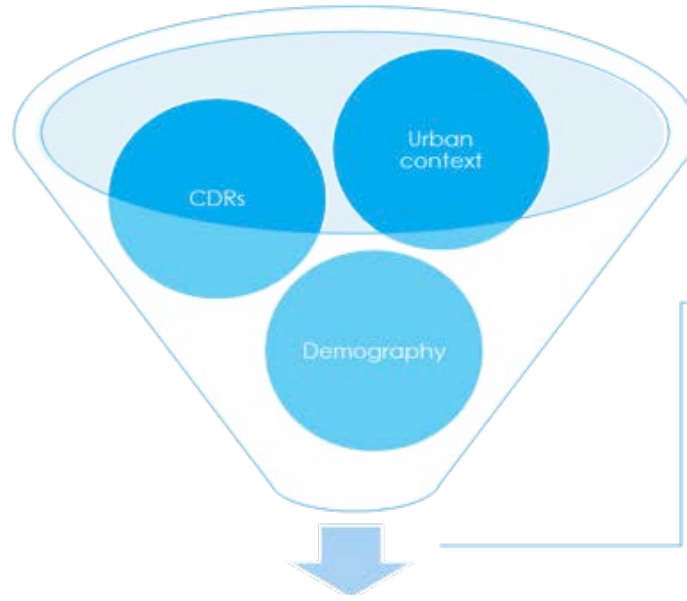
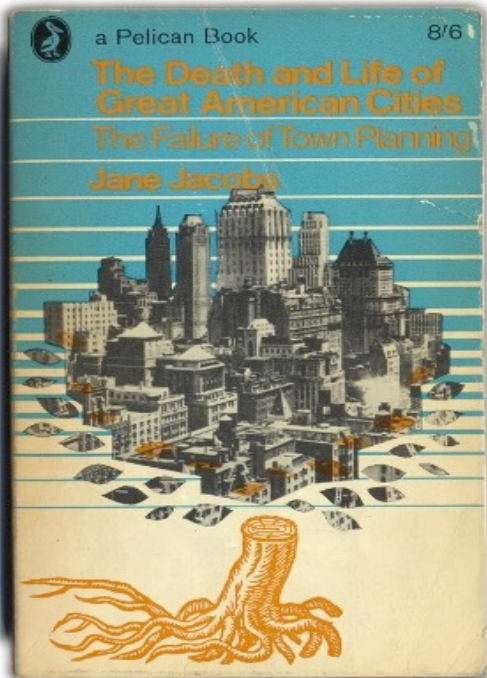
Criminality?

Project #2: Urban Safety and Vitality

Inside Out: to understand crime mechanisms look at urban fabric first

Marco De Nadai^{*1,2}, Emmanuel Letouzé^{†4}, Marta C. González^{‡3} and Bruno Lepri^{§2}

¹University of Trento. ²FBK. ³MIT. ⁴Data-Pop Alliance



Criminality?

“Preliminary findings indicate that Jacobs' diversity conditions are a better predictor of homicides and robberies than socio-economic conditions.”

BUT police records are probably biased; do not take at face value!....*

*Big Risk: Bias and Unrepresentativeness

The ABCDE of Big Data: Assessing Biases in Call-detail records for Development Estimates

Gabriel Pestre*
Data-Pop Alliance

Emmanuel Letouzé†
Data-Pop Alliance

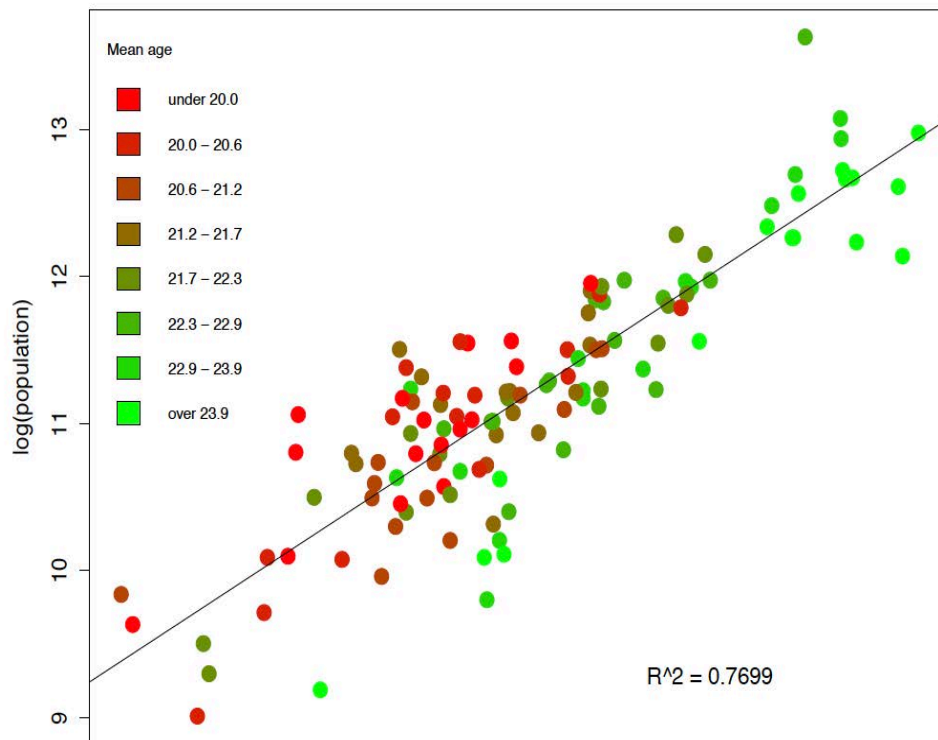
Emilio Zagheni‡
University of Washington, Seattle

March 15, 2016

Abstract

This article contributes to improving our understanding of biases in estimates of demographic indicators, in the developing world, based on Call Detail Records (CDRs). CDRs represent an important and largely untapped source of data for the developing world. However, they are not representative of the underlying population. We combine CDRs and Census data for Senegal in 2013 to evaluate biases related to estimates of population density. We show that: (i) there are systematic relationships between cellphone use and socio-economic and geographic characteristics that can be leveraged to improve estimates of population density; (ii) when no 'ground truth' data is available, a difference-in-difference approach can be used to reduce bias and infer relative changes over time in population size at the subnational level; (iii) indicators of development, including urbanization and internal, circular, and temporary migration, can be monitored by integrating Census data and CDRs. The paper is intended to offer a methodological contribution and examples of applications related to combining new and traditional data sources to improve our ability to monitor development indicators over time and space.

Preliminary Draft prepared for submission to the Annual World Bank Conference on Development Economics - 2016



$$\log(P_i) = k \log(U_i) + bias_i + \varepsilon_i$$

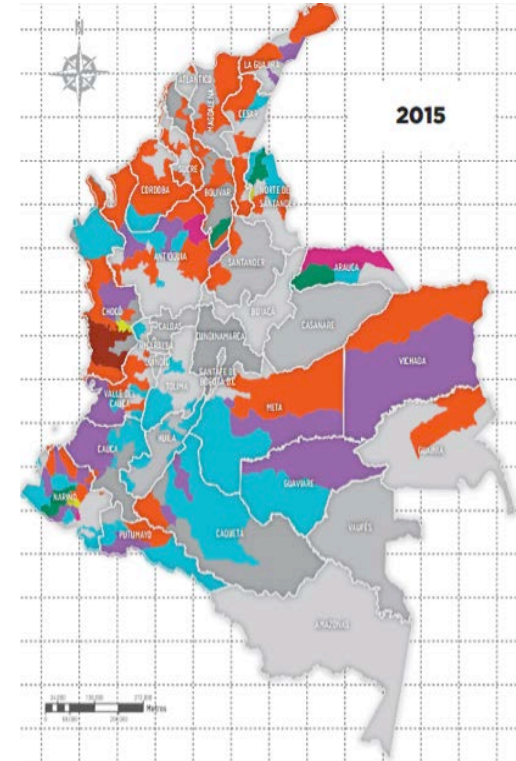
$$bias_i = f(\text{mobile phone penetration}_i)$$

$$Y_i = \beta_0 + \beta_1 \text{mobile phone penetration}_i + \varepsilon_i$$

Project #2: Urban Safety and Vitality

Next: \$300k research program with IDB

- In partnership with the city governments of Bogotá, Cartagena, Cali, Medellin, and Santa Marta
- Key (tentative) research questions identified at Media Lab workshop in March 2016:
 - *How will the implementation of the peace agreement affect people's trust in the police?*
 - *How / why are police records biased?*
 - *Which are the key determinants of crime?*
 - *What policy interventions may reduce criminality?*

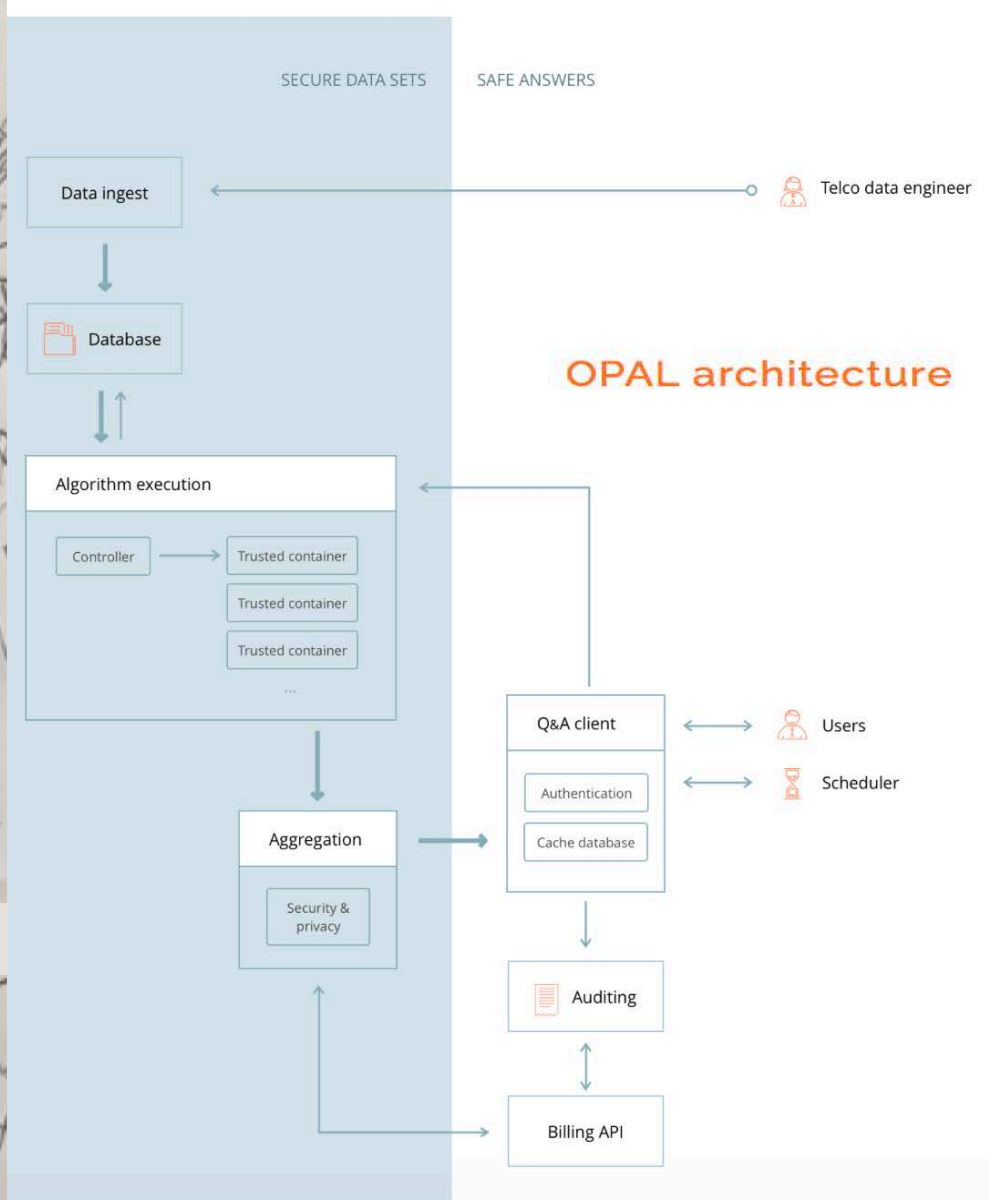


Project #3: Open Algorithms (OPAL)



OPAL
bring the code to the data





Project #3: Open Algorithms (OPAL)

Open algorithms: A new paradigm for using private data for social good

By Thomas Roca, Emmanuel Letouzé | 18 July 2016



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



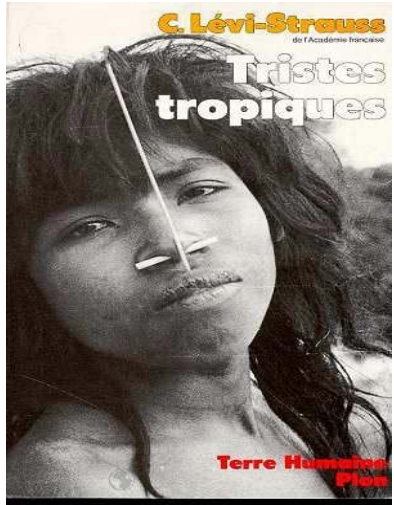
Project #4: Global Data Literacy Program

October 2015

DATA-POP ALLIANCE
WHITE PAPER SERIES

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

We conceptualize data literacy as literacy in the age of data, defined as “the desire and ability to constructively engage in society through or about data”.



“[In 19th Century Western Europe], the fight against illiteracy [was] indistinguishable from the increased control exerted over the individual citizen by the holders of power.”

Building Literacy for the Data Generation

December 18, 2015

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



Ed...
July 2012
MANU

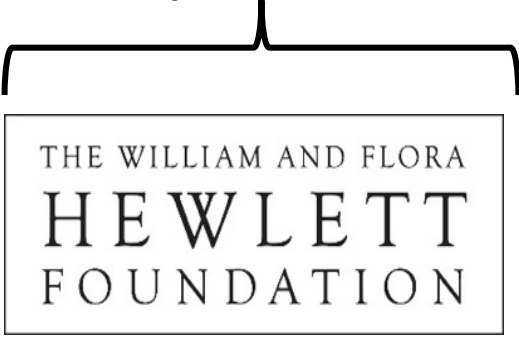
Project #4: Global Data Literacy Program

Professional training workshops on Big Data and Sustainable development in Colombia and 6 other countries over 2016-17



Toolkit + Online curated library in English, Spanish, and French

Main funders/ partners



Selected technical partners



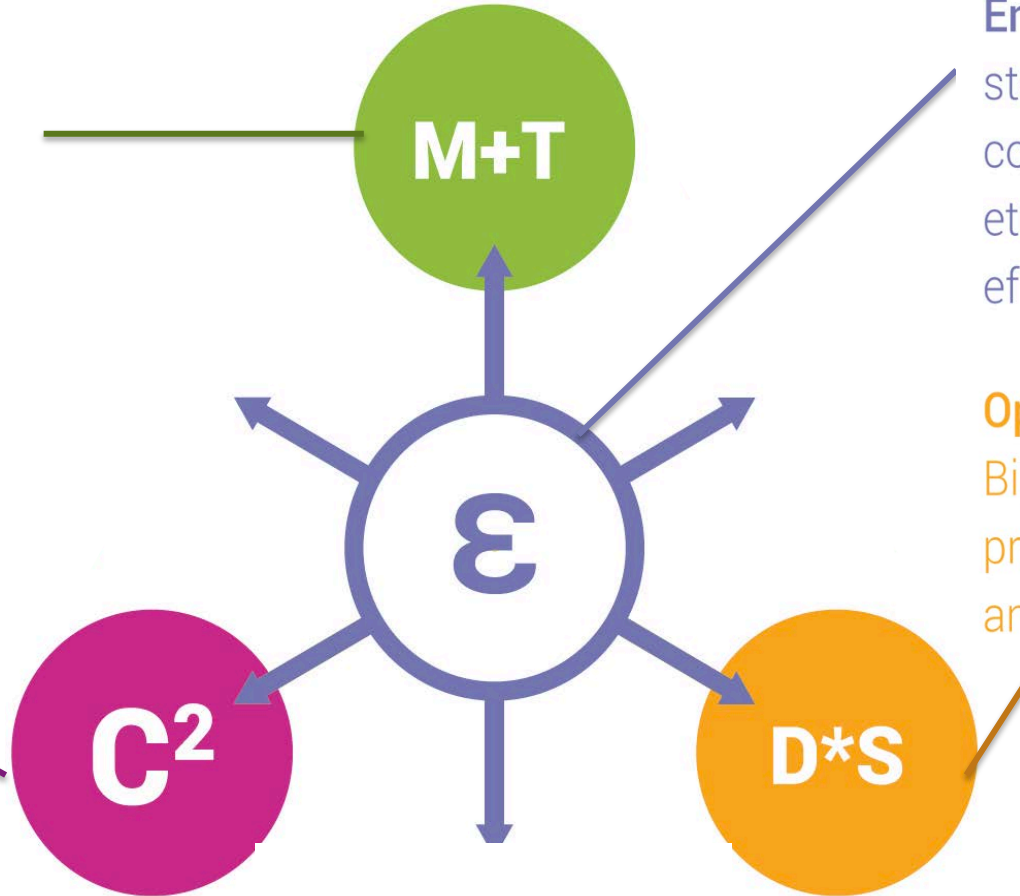
Selected target audience

- UN Officials
- Gov officials
- Official Statisticians
- Journalists
- NGO/CSO Staff
- Academics/

Testing Building Blocks of 'Big 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

Project #5: Colombia's Big Data strategy



DNP Departamento Nacional de Planeación



TODOS POR UN NUEVO PAÍS
PAZ EQUIDAD EDUCACIÓN



"We are at the beginning of what I call *The Industrial Revolution of Data.*"
Joseph Hellerstein, November 19, 2008

Computational Social Science
A field is emerging that leverages the capacity to collect and analyze data at a scale that may reveal patterns of individual and group behaviors. (*Science*, 2009)

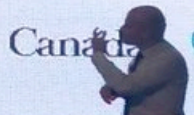


The Economist Data and development
Off the map



Africa's statistical tragedy

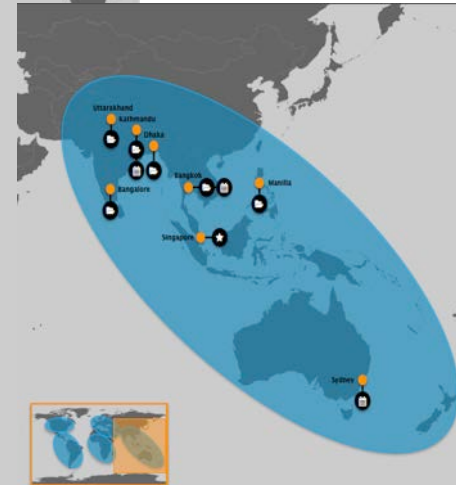
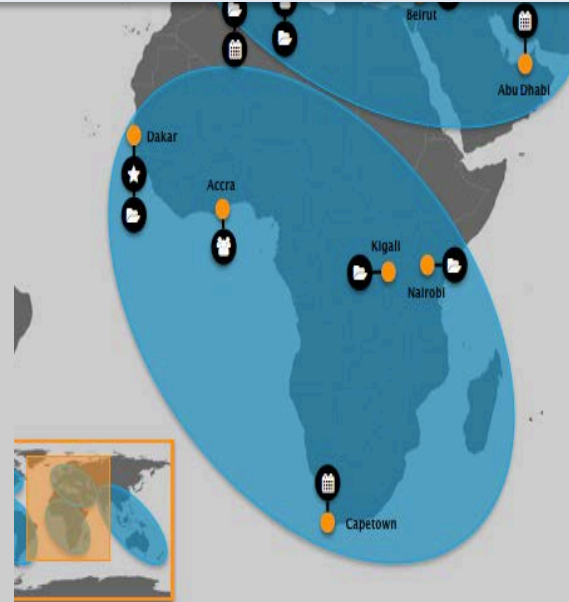
Rich countries are deluged with data; developing ones are suffering from drought
Nov 15th 2014 | NEW YORK AND LONDON | From the print edition



Announcement of DNP-Data-Pop Alliance agreement to support Colombia's National Big Data Strategy at Andicom, Cartagena, September 2016

Plan: Experiencing Locally, Globally

Plan: co-design and deploy connected research and development activities anchored in major cities around the world as part of regional Data Spaces—starting in Bogotá since 2013, then Dakar, Nairobi, Manilla ...to test hypotheses and methods on how data can be leveraged to build fairer and safer societies



Concluding Thoughts

1. I am convinced we can and must, **in our lifetime and for our children, leverage data to build fairer and safer societies**
2. **'We' know very little about to do this at scale, systematically, sustainably; it will be much more disruptive than most think; even less little is done to try out**
3. I want to **dedicate my future career to (1) understanding and (2) implementing** ways to do this
4. I think I am **very well positioned to do this successfully; and would be in a much better position at the Media Lab** (and I also want to use cartoons more, which is the only place I could)
5. I would appreciate **any feedback/questions/expressions of interest to be involved** in these plans.

Thank you
eletouze@mit.edu

Appendix



“Your test results came back—I’m especially worried about your credit score.”