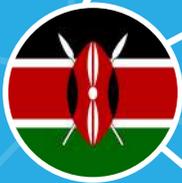




DATA-POP ALLIANCE



KENYA

— LEVERAGING —
BIG DATA
— for —
SUSTAINABLE
DEVELOPMENT

Professional Training
Workshop

June 27-29, 2017
United Nations Office at Nairobi (UNON)
Nairobi, Kenya

PARTICIPANT GUIDEBOOK

ABOUT US

Promoting a people-centered Big Data revolution



About Data-Pop Alliance

Data-Pop Alliance is a global coalition on Big Data and development created in 2014 by the Harvard Humanitarian Initiative (HHI), MIT Media Lab and the Overseas Development Institute (ODI), joined in February 2016 by the Flowminder Foundation as a fourth core member, to “promote a people-centered Big Data revolution.” Data-Pop Alliance started with seed funding from the Rockefeller Foundation and has since then been supported by a range of bilateral and multilateral donors, philanthropic foundations, and private companies.

Data-Pop Alliance brings together researchers, experts, practitioners, and activists working around the globe across different disciplines and industries around a common vision: making Big Data a force of positive social change in the 21st Century. To that end, Data-Pop Alliance focuses on the applications and implications of Big Data for development and seeks to ensure that Big Data contributes to improving decision-making processes and fostering citizen empowerment in ways that avoid elite capture, widening inequities, and the dehumanization of public policies.

Data-Pop Alliance works through 3 main modalities, deployed on the ground in various regions of world: (1) collaborative research, with both empirical and ‘White Papers’, etc.; (2) capacity building, through trainings, workshops, Data Expeditions, and more; and (3) policy and community engagement, including through our involvement in the Global Partnership for Sustainable Development Data, the UN World Data Forum and in-country partnerships

and activities. Our core domains of expertise and focus include official statistics and governance; urban dynamics; demographic and economic processes; peacebuilding and social cohesion; climate change and resilience; and data literacy and ethics.

Data-Pop Alliance functions as a distributed network with its headquarters and core team hosted at ThoughtWorks in New York City, a Bogotá-based team, and directors in New York City, Cambridge, London, and Geneva. In addition, Data-Pop Alliance relies on the intellectual resources of its inner circle of close to 30 Research Affiliates, experts and scholars located in more than 15 countries, and a network of more than 30 technical partners.



About the United Nations System Staff College (UNSSC)

With its Headquarters in Turin (Italy) and its second Campus in Bonn (Germany), the United Nations System Staff College (UNSSC) is a centre of excellence for training and knowledge management within the UN system. Each year, we offer learning initiatives that reach over 10,000 beneficiaries including residential courses, distance-learning, strategic exchanges and seminars.

We also provide learning interventions to increase technical and substantive knowledge spanning the three pillars of the UN’s work: peace and security, human rights and development. Since 2015, all of the Staff College’s learning and training activities are geared towards enabling the UN system and its partners to address the challenges posed by the adoption of Agenda 2030 for Sustainable Development.

To this end, the UNSSC, through the establishment of the Knowledge Centre for Sustainable Development on January 2016 in Bonn, facilitates comprehensive learning, training, and knowledge management needs of UN staff and partners in the context of the 2030 Agenda as well as the Paris Agreement under the United Nations Framework Convention on Climate Change.

The Knowledge Centre for Sustainable Development also serves as a catalyst and convener prompting dialogue and knowledge sharing on issues relevant to the vision and mission of the United Nations between UN staff and a diverse set of stakeholders from academia, the private sector and civil society.

Trainings developed with financial support from



WELCOME

Dear Participants,

Thank you for registering for the inaugural workshop of the Global Professional Training Program on Big Data and Sustainable Development, and we look forward to welcoming you in Nairobi, Kenya, on June 27 to 29!

During this three-day workshop, we will provide you with the skills and resources needed to navigate and leverage Big Data in and through your professional environments. The workshop will include expert presentations, facilitated technical tutorials, group work and a Data Expedition, supported by dedicated materials and datasets available off and online, covering both applications and implications of Big Data for development policy and program—including conceptual, technical, methodological, political, and ethical.

We approach 'Big Data' not simply as large datasets to be extracted and refined to recommend products or predict GDP, but as a socio-technological phenomenon of historical significance, shaped by the "3 Cs" of Big Data—at its core large volumes of passively emitted digital 'crumbs', ever more powerful analytical, technological and human capacities, and diverse communities of citizen-data generators, analysts and users responding to and shaping different values and incentives.

The workshop is structured around four key dimensions, or "building blocks," for developing practical data literacy skills for practitioners and policy-makers to build inclusive Big Data-driven innovation projects, policies and partnerships. These building blocks include what we have termed C² for Contexts and Concepts, M+T for Methods and Tools, D*S for Design and Strategy, and, last but certainly not least, as a primary lens rather than an afterthought, ε for Ethics and Engagement. We also hope the personal connections you make during this workshop will enable you to build lasting collaborations and future partnerships.

This guidebook will help you to become more familiar with the concepts we will be discussing, and will also allow you to think more about your personal goals for the workshop and develop questions you would like answered. It also includes important information on travel, lodging, the venue and technological requirements.

Please let us know if you have any questions by contacting us at trainings@datapopalliance.org.

We are glad and grateful you are able to join us for this workshop, and again welcome!

A handwritten signature in black ink that reads 'Emmanuel Letouzé'.

Emmanuel Letouzé, PhD

Director & Co-Founder | Data-Pop Alliance
Visiting Scholar | MIT Media Lab

TABLE OF CONTENTS

- 2** About the Training Program
- 3** Overview and Learning Objectives
- 4** Workshop Agenda
- 6** Speakers
- 15** Logistical Information
- 18** Pre-Workshop Preparation
- 20** Big Data Literacy Primer
- 22** Timeline of The Data Revolution
- 23** The 3 C's of Big Data
- 25** Contexts & Concepts
- 27** Methods & Tools
- 29** Design & Strategy
- 31** Ethics & Engagement
- 33** Glossary of Terms

Trust
me



i'm
a data scientist

ABOUT THE TRAINING PROGRAM

Why a global professional training program in Big Data and development?

This training program is designed specifically for professionals who have a role in defining ways that data can support sustainable development—through policy, public services and programs, technology products, and more. By bringing together professionals who come at the question of “What is the significance of Big Data” from diverse perspectives, the training will create a space for these leaders to define answers that are directly relevant and applicable in their own contexts.

The training aims to ensure that these leaders are more equipped to work with data and data scientists and other stakeholders to design and execute data-driven interventions towards sustainable development projects.

Ultimately, the training program aims to help participants:

- 1 Develop an understanding and practical skills of Big Data for society and in their professional environments;
- 2 Constructively engage in discussions on the applications and implications of Big Data for development;
- 3 Understand and describe a range of methods, tools and analyses relevant to leveraging data for sustainable development through hands-on, technical sessions;
- 4 Consider the responsible use of Big Data within their own organizations and through partnerships;
- 5 Acquire practical tools, resources, and support network to continue growing their capacity to leverage Big Data for social good.

OVERVIEW AND LEARNING OBJECTIVES

Corresponding to what we consider the four building blocks of data literacy, the workshop is structured around four key dimensions, or “building blocks” for developing practical data literacy skills for practitioners and policy-makers to build inclusive Big Data-driven innovation projects, policies and partnerships. This workshop incorporates a blend of teaching formats, including traditional instructor-led teaching, keynote speakers, and interactive and hands-on technical sessions.

The following summary highlights the key learning points related to each building block:

C² Contexts & Concepts

- 1** Decode key terms and buzzwords in the Big Data and development landscape
- 2** Discuss Big Data within the broader political context of the post-2015 sustainable development framework and data for social good
- 3** Translate development problems into specific data objectives

M+T Methods & Tools

- 4** Understand existing methods and tools used to leverage Big Data
- 5** Assess data representativeness, biases and insights within Big Data-driven approaches and methods
- 6** Identify applicable tools by assessing the value add of Big Data for specific development problems

D*S Design & Strategy

- 7** Identify individual and organizational objectives towards a Big Data strategy
- 8** Understand how to operationalize Big Data as projects, partnerships and policies
- 9** Recognize individual and organizational next steps towards Big Data applications

E Ethics & Engagement

- 10** Identify models for prioritizing inclusivity, transparency and accountability in data public-private-people partnerships
- 11** Articulate and assess ethical, privacy and legal implications of Big Data applications
- 12** Understand key principles for effective data communication and story-telling

WORKSHOP AGENDA

DAY 1 | JUNE 27



Contexts and Concepts: Understanding the foundations of the Big Data for development ecosystem

8:30 - 9 AM REGISTRATION

9 AM - 1 PM MORNING SESSIONS

- | | |
|---|---|
| <p>1 Kickoff - What to expect and your questions on the 2030 agenda</p> <p>2 Interactive Lecture - Monitoring and promoting the SDGs in the age of Big Data</p> | <p>3 Panel Discussion - Case studies on Big Data and the 2030 agenda for sustainable development</p> |
|---|---|

1 - 2 PM LUNCH

2 - 5.30 PM AFTERNOON SESSIONS

- | | |
|---|---|
| <p>1 Designing Big Data Projects - Mapping Big Data and development ecosystems</p> <p>2 Big Picture - Social Physics: How social networks can make us smarter (via video)</p> | <p>3 Designing Big Data Projects - Breaking big development problems into data questions</p> |
|---|---|

DAY 2 | JUNE 28



Methods & Tools: Identifying Big Data methods and tools to translate sustainable development challenges to specific data questions

9 AM - 1 PM MORNING SESSIONS

- | | |
|---|--|
| <p>1 Warm Up - Programming icebreaker</p> <p>2 Panel Discussion - Data to action: movements and tools</p> | <p>3 Technical Session - Part I. Data visualization: when statistics meets web & data science</p> |
|---|--|

1 - 2 PM LUNCH

2 - 5.30 PM AFTERNOON SESSIONS

- | | |
|--|--|
| <p>1 Technical Session - Part II. Data visualization: when statistics meets web & data science</p> <p>2 Interactive Lecture - Citizen generated data with community journalism</p> | <p>3 Designing Big Data Projects - Mapping Big Data methods and tools</p> |
|--|--|

5.30 PM HAPPY HOUR

WORKSHOP AGENDA

DAY 3 | JUNE 29



Design & Strategy: Operationalizing Big Data projects, partnerships and policies by engaging stakeholders and communities through ethical practices and effective story-telling

9 AM - 1 PM MORNING SESSIONS

- | | |
|---|--|
| 1 Technical Session - Data storytelling: telling the story you want to tell with data | 3 Designing Big Data Projects - Project archetypes and lessons learned |
| 2 Interactive Lecture - Inclusion, engagement and collaboration in data and development initiatives | 4 Interactive Lecture - Legal and regulatory data frameworks in Kenya |

1 - 2 PM LUNCH

2 - 5.30 PM AFTERNOON SESSIONS

- | | |
|---|---|
| 1 Designing Big Data Projects - Data ethics lab | 3 Building a Community of Practice - Reflections, feedback and next steps |
| 2 Designing Big Data Projects - Presentations | |

SPEAKERS



Madame Sahle-Work Zewde | @UnonDg
UNON DG

Madame Sahle-Work Zewde of Ethiopia was appointed as the Director-General of the United Nations Office at Nairobi (UNON) by the Secretary-General, Mr. Ban Ki-moon, on 11 March 2011. Madame Zewde is the first person to be appointed to the newly created position as head of the Organization's Nairobi Office, at the Under-Secretary-General level.

Madame Zewde brings to this position more than two decades of experience at the national, regional and international levels. Prior to taking up her functions as

Director-General of the United Nations Office at Nairobi, Madame Zewde served as Special Representative and Head of the United Nations Integrated Peacebuilding Office in the Central African Republic.

A career diplomat, Madame Zewde served as Permanent Representative of Ethiopia to the African Union and the United Nations Economic Commission for Africa from 2006 to 2009. She also served as Director-General for African Affairs in the Ministry of Foreign Affairs of Ethiopia from 2006-2009. Prior to that appointment, Madame Zewde was Ambassador to France, Permanent Representative to the United Nations Educational, Scientific and Cultural Organization (UNESCO) and accredited to Tunisia and Morocco (2002 until 2006). She also served as Ambassador to Djibouti and Permanent Representative to the Intergovernmental Authority for Development (IGAD) from 1993 to 2002, and Ambassador to Senegal, with accreditation to Mali, Cape Verde, Guinea-Bissau, Gambia and Guinea from 1989 to 1993.

Ms. Zewde is a natural sciences graduate of the University of Montpellier, France. She is fluent in Amharic, French and English. Born on 21 February 1950, Madame Zewde has two sons.

Davis Adieno | @DavisAdieno
Civicus

Davis works at CIVICUS World Alliance as Senior Advisor, Sustainable Development. He advises, oversees and coordinates organisational initiatives on sustainable development, including identifying and pursuing partnerships and fundraising opportunities for these. He designs organisational positions and messaging on sustainable development in consultation with CIVICUS' membership and other key stakeholders. He plays a leadership role in the strategic direction and delivery of DataShift



SPEAKERS

(an initiative of CIVICUS), and has extensive experience in data, citizen engagement, and policy engagement. He joined CIVICUS in 2016 from Development Initiatives where he worked as Senior Manager Strategic Partnerships and Data Use, based at the Africa Hub in Nairobi Kenya. He champions the harnessing of the data revolution for effective sustainable development at various levels – by supporting and bringing together parliamentarians, policy makers, civil servants, technology experts, civil society organisations, and community level organisations to develop inclusive information ecosystems that support data sharing and use. He holds a Masters in Communication Studies from the University of Nairobi in Kenya.



Bill Anderson | @devinitorg
Development Initiatives

Bill Anderson is an Information Architect with Development Initiatives, leading work on joined-up data standards and the role of sub-national data in local decision-making. He is also Technical Lead for the International Aid Transparency Initiative. He leads a team of data scientists in DI and provides oversight of the development of systems underpinning DI's collection, storage and analysis of data and production of information. He is a South African who served the African National Congress for fifteen years as an intelligence analyst in the struggle against apartheid.

Ugo Caruso | @UgCaruso
UNSSC

Ugo Caruso currently serves as the Course Coordinator for the United Nations System Staff College Knowledge Centre for Sustainable Development. Prior to joining the UN System Staff College in Turin in 2008, Ugo worked for the European Centre for Minority Issues, the Council of Europe Roma and Migration Department and was lecturer and researcher on human and minority rights at the Universities of Southern Denmark (Odense campus) and Johann Wolfgang Goethe University of Frankfurt am Main (Germany).



He was also visiting researcher to the South Asian Association for Regional Co-operation (Kathmandu, Nepal), the Max Planck Institute for Comparative Public Law (Heidelberg, Germany) and International Law and the Walther Schücking Institute for International Law (Kiel, Germany). He is also author and editor of several books and articles on human

SPEAKERS

and minority rights issues, and UN-regional organizations cooperation in conflict-related scenarios. He holds a Ph.D. in Public International Law from the Johann Wolfgang Goethe University of Frankfurt am Main.



Nixon M. Gecheo | @nmageka
Communications Authority of Kenya

Nixon Mageka Gecheo is the immediate former ICT and Media Advisor to the Cabinet Secretary of Agriculture. He is a governance and ICT4D consultant and currently a Member of the Universal Service Advisory Council of the Communication Authority of Kenya, Advising on the use of Universal Service Fund in Spreading ICTs to the Underserved and un-served areas of the Republic of Kenya. Nixon also serves as a council member of the Turkana University College.

Prior to being a consultant, Mageka worked for 10 years in Public Sector mainly in Public sector reforms and the role of information technology in organizations. Mageka worked as a Chief ICT officer for the Commission on Administrative Justice also known as the Ombudsman.

Al Kags | @alkags
Open Institute

Al Kags is the founder of the Open Institute, an organisation based in Nairobi, Kenya that works with governments and civil society organisations to promote Open Government and Citizen Participation. Primarily, Al works across sub-saharan Africa but contributes to global open government movements. He is currently also a member of the Global Open Data Working Group and the Global Open Data Initiative. He is a co-chair of the Kenya OGP Working Group and was the chair of the Kenya Open Data Taskforce and the Kenya Open Government Initiative.



Al is a Mandela Washington Fellow (2014) and was recognised as a New Generation African Leader (2013). In 2016 Al Kags was awarded the Open Data Champion award by the Kenya ICT Authority.

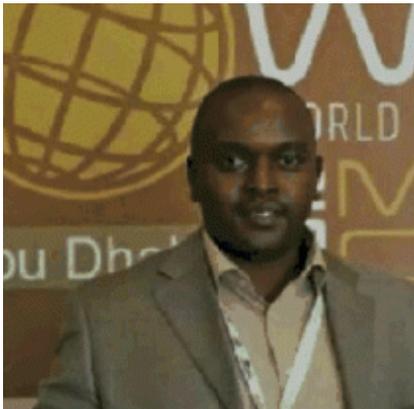
SPEAKERS



Emmanuel Letouzé | @ManuLetouze
Data-Pop Alliance

Emmanuel Letouzé is the Director and co-Founder of Data-Pop Alliance, a coalition on Big Data and development co-created in 2013 by the Harvard Humanitarian Initiative, MIT Media Lab, Overseas Development Institute, joined in 2016 by the Flowminder Foundation as its 4th core member. He is a Visiting Scholar at MIT Media Lab, a Research Affiliate at HHI and a Research Associate at ODI. He is the author of UN Global Pulse’s White Paper “Big Data for Development” (2012) and of the 2013 and 2014

OECD Fragiles States reports. His research and work focus on Big Data’s application and implications for official statistics, poverty and inequality, conflict, crime, and fragility, climate change, vulnerability and resilience, and human rights, ethics, and politics. He worked as a Development Economist for UNDP in New York from 2006-09 on fiscal policy, post-conflict economic recovery and migration, and between 2000-04 in Hanoi, Vietnam, for the French Ministry of Finance as a technical assistant in public finance and official statistics. He holds a BA in Political Science and an MA in Economic Demography from Sciences Po Paris, an MA from Columbia University School of International and Public Affairs, where he was a Fulbright Fellow, and a PhD from the University of California, Berkeley. He also a political cartoonist for various publications and media as ‘Manu’.



Silas Macharia | @SilasMacharia
IBM Kenya

Silas Macharia (silaska@ke.ibm.com) Silas has over 15 years experience working in the ICT Industry across Africa leading in Offering/Product Software consulting engagement to enable customers meet business objectives. He has led the first implementation of Research Based Open Data Platform using API’s(catalogue and Marketplace) on the continent. Silas has been instrumental in naturalizing technology to capture African Context.

Silas spends significant amounts of time presenting to clients enabling them to deliver exceptions customer experience to the end customer. Silas previously worked at a Technology Company that was a winner of Two Global Awards in the space of supply chain management. He has significant expertise in the areas of enterprise Cloud ,mobile technologies, Open Data, Disruptions and Data Monetization technology having taken on diverse roles through his career with I.T. Space and Beyond.

SPEAKERS



Joseph Marindi | @marindijm
Data Manager, Humanitarian Data Exchange (HDX)

Joseph Marindi is an ICT professional with a wide range of experience in Telecommunications and NGO sectors. Currently he is the Humanitarian Data Exchange (HDX) data manager with the United Nations office for coordination of Humanitarian Affairs (UN-OCHA). He has previously served as an Information systems manager with Innovations for poverty action and as an ICT projects coordinator with Orange Telkom Kenya. He holds an MBA (strategic management) from the University of Nairobi and a Bachelor's Degree in

Computer science. He is passionate about open data and Evidence based decision making.

Ana Lucía Martínez | @analutweeting
Data-Pop Alliance

Ana Lucía holds a B.A. in Industrial Design and a Post-grad degree in Economic Evaluation of Projects from Universidad de Los Andes in Bogotá. She has a strong background in the generation of strategies to improve organizational performance. After finishing her PgD she worked for 4 years at the National Statistical Office in Colombia, where she held different positions overseeing the implementation of various projects. Ana Lucía's interests include data visualization and design for development.



Zacharia Muindi | @muindre
Map Kibera Trust



Zach works with Map Kibera as the lead mapper, and have worked there from when it began in 2009 up to date. He was part of the first team that mapped out Kibera and developed a digital map of the area. Zach has also participated in different mapping projects on health, education, watsan and security. He has also trained youths from other parts of the countries and universities on how they can use mapping as a digital tool to highlight issues that affect them.

SPEAKERS



Joshua Ogure | @Joshculture
Map Kibera Trust

Joshua Ogure famously known as Josh, is the project manager for Map Kibera Trust (www.mapkibera.org) he is also the team leader for Kibera News Network (www.kiberanewsnetwork.org) a citizen journalism program in Kibera Slum Nairobi Kenya. In his community, Josh strives to bring about positive social change by using the lens of his camera to engage and document what happens in and around his neighbourhood. Josh has a wide range of experience working in the informal settlements of Nairobi including Mathare and Mukuru slums. As a project manager, in 2014/2015

Josh lead his team towards a successful Pilot project dubbed Open Schools Kenya (www.openschoolskenya.org) that put all Kibera schools on the map and developed a website for them. As he strive to make the invisible visible using data and journalism for advocacy. Josh also uses his platforms to explore, investigate and give people the right information as he also holds the leaders accountable. Josh recently won and attended a four months IREX's Community Solutions Program in the United States from where he was awarded a certificate for demonstrating an outstanding leadership on peace and conflict resolution by George Mason University and the United States Department of States. November 2016 , Josh on behalf of Map Kibera Trust won the Kenya Open Data award under the social impact category in November 2016.

Sanda Ojiambo | @SandaOjiambo
Safaricom

Sanda Ojiambo is the Head of Corporate Responsibility at Safaricom, where she manages the Social Innovation, Sustainability, Environmental Management, Social Policy and the Corporate Social Investment portfolios. Sanda has a Master's degree in Public Policy and Development Economics. She has completed training in Executive Leadership at Strathmore Business School in Kenya and Harvard Business School in the USA.



She previously worked as the Director of Programmes at the Africa Regional Office of the International Planned Parenthood Federation, where she provided technical and strategic leadership for the development and implementation of women's health programmes in over 40 countries in sub-Saharan Africa. Sanda has also worked at CARE International, UNDP Somalia and Planned Parenthood Federation of America - International; where she managed a range of development projects including community based health and education interventions; NGO capacity building, women's empowerment

SPEAKERS

and women's health. She is a Director of the Kenya Investment Authority (KenInvest), a Kenyan State Corporation; and is a board member of the Gender Violence Recovery Centre, a Kenyan NGO. Sanda is passionate about contributing to Kenya's development agenda and the evolving and growing role of technology in providing socio-economic development solutions. She is an avid traveler and recently climbed Mount Kilimanjaro.



Sandy Pentland | @alex_pentland

MIT

Professor Alex 'Sandy' Pentland directs MIT's Human Dynamics Laboratory and the MIT Media Lab Entrepreneurship Program, co-leads the World Economic Forum Big Data and Personal Data initiatives, and is a member of the Advisory Boards for Nissan, Motorola Mobility, Google, Telefonica, and a variety of start-up firms.

He is the Academic Director of Data-Pop Alliance. He has previously helped create and direct MIT's Media Laboratory, the Media Lab Asia laboratories at the

Indian Institutes of Technology, and Strong Hospital's Center for Future Health.

Sandy's research group and entrepreneurship program have spun off more than 30 companies to date, three of which are publicly listed and several that serve millions of poor in Africa, South Asia, and Latin America. Recent spin-offs have been featured in publications such as the Economist and the New York Times, as well as winning a variety of prizes from international development organizations.

Julie Ricard | @ricardjulie

Data-Pop Alliance

Julie Ricard is the Research and Communications Officer at Data-Pop Alliance. She graduated with two Masters degrees (1) in International Development by the Paris School of International Affairs (Sciences Po Paris) and (2) in Gender Studies by Université Paris Diderot (Paris VII). She also holds a B.A. in Latin American Studies (Sciences Po Paris). Prior to working at Data-Pop Alliance, she worked at the French Agency for Development, supporting the implementation of the gender strategy by doing research and developing capacity-building knowledge products. Most recently, as an International Data Responsibility Group fellow, she supported research about the use of ITs in migration and refugee camps. At Data-Pop Alliance, Julie's key responsibilities include project management, research and internal and external communications.



SPEAKERS



Thomas Roca | @Thomas_Roca
French Development Agency

Thomas Roca is a researcher and statistician at the French Development Agency (AFD) Research Directorate. Thomas is developing AFD's research program covering well-being, Human development, alternative welfare indicators & Data for Development. Thomas' field of work also covers data visualization and programming. Within this research program, partnerships have been built with Data-Pop Alliance, UNDP, the World Bank, UNU-MERIT. In 2015 Thomas was a visiting research fellow at the Human Development Report Office, UNDP (NY). As a

statistician, Thomas was in charge of the Institutional Profiles Database 2012 (IPD), a global perception survey, probing institutions and governance quality. Before joining AFD in 2011, Thomas taught Economics in Hungary, Vietnam, and Kurdistan as a lecturer for Université de Picardie. During his PhD, probing governance measurement, Thomas interned twice at UNDP: first in Morocco within the governance program (2006), then in New York at the Human Development Report Office (2010).

Natalie Shoup | @natshoup
Data-Pop Alliance



Natalie graduated with a BSE in Operations Research and Financial Engineering from Princeton University. Prior to joining Data-Pop Alliance, Natalie worked with a Big Data Analytics consultancy firm based in New York. Natalie has spent a great deal of time in Latin America. After graduation she served as a Princeton in Latin America Fellow in partnership with Developing Minds Foundation, working with NGOs focusing on education in Brazil and Colombia. As Data-Pop Alliance's Program Manager, Natalie's key responsibilities include overseeing project design and delivery, conducting and coordinating research and developing partnerships.

SPEAKERS



Skyler Speakman | @PhonesDrones
IBM Research -- Africa

Skyler Speakman is a Research Scientist at IBM Research -- Africa. His projects use data science to impact the lives of millions of people on the continent. Skyler completed his Ph.D. in Information Systems at Carnegie Mellon University as well as a M.S. in Machine Learning. He lives in Nairobi, Kenya with his wife and two young sons.

Patrick Vinck - @developmentdata
Harvard University

Patrick Vinck, PhD, is the co-Founder and co-Director of Data-Pop Alliance; he is the Harvard Humanitarian Initiative's Director of Research, an Assistant Professor at the Harvard Medical School and Harvard T.H. Chan School of Public Health, and Lead Investigator at the Brigham and Women's Hospital. His current research examines resilience, peacebuilding, and social cohesion in conflicts and disaster settings, as well as the ethics of data and technology in the field. He is the co-Founder and Director of KoBoToolbox a data collection service. Prior to joining HHI in 2011, he founded the program at the University of California Berkeley's Human Rights Center.



LOGISTICAL INFORMATION

Venue

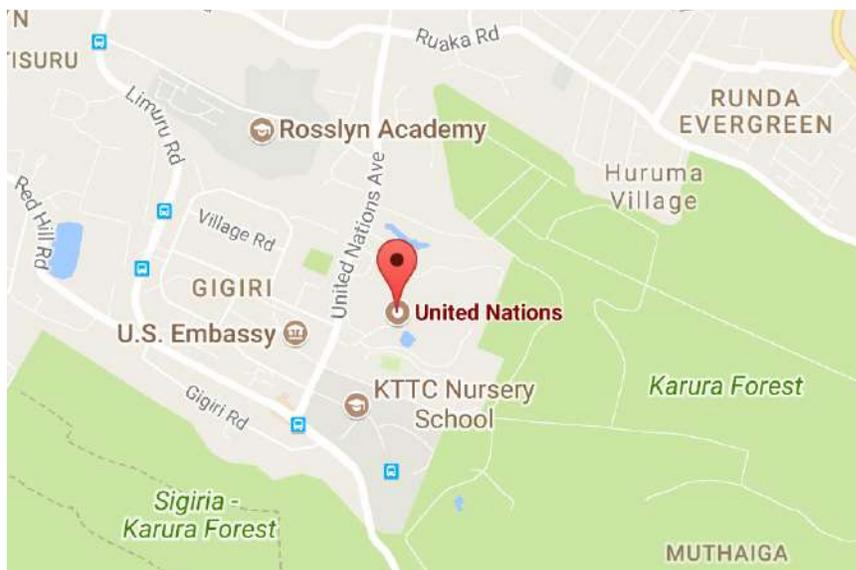
The course will be held at the UN office at Nairobi (UNON)

Gigiri Complex, P.O. Box 30552 (00100), Nairobi, Kenya.

Room: Conference Room 14

Location: Rooftop (from the main security lobby)

In order to access UNON, please make sure to have a valid national ID card or passport.



Food and Drink

Coffee breaks will be provided each day. Lunch will be at the responsibility of participants.

Refreshments will be provided during the Happy Hour event on June 28.

Weather and Attire

The month of June is characterized by gradually falling daily high temperatures, with daily highs around 23°C throughout the month, exceeding 26°C or dropping below 20°C only one day in ten.

Course attire is business casual.

Language

The workshop and course materials will be in English. To get the most value from the workshop, participants should be comfortable discussing issues around data and their work in English. Please contact Data-Pop

if you have any questions or concerns related to language.

Course material

Secretarial services will be made available at the course. The course workbook with the documentation for the course will be available to the participants on the first day of the programme. Please note that UNSSC applies a green policy, therefore we will limit the use of paper during the course. The course materials will be made available on our social learning platform (CLANED), access to the platform will be given to registered participants in advance of the course.

Computers

Participants are expected to bring a laptop to be able to participate fully in the workshop sessions, especially on Day 2. Electrical outlets could be limited, so please fully charge laptops and phones before arriving at the venue if possible.

LOGISTICAL INFORMATION

Twitter # for the course
#Data4SDcourse

Contact information

If you have any questions, please contact one of the members of our team:

Natalie Shoup or Julie Ricard
Data-Pop Alliance
Email: trainings@datapopalliance.org

Ugo Caruso
UNSSC Knowledge Centre for Sustainable Development
Email: u.caruso@unssc.org
Tel: +49 228 815 2655

Fatma Gatabaki
UNON
Email: Fatma.Gatabaki@unon.org
Tel: +254 733 874133

Accommodation

Participants are responsible for their own accommodation. There are many recommended hotels and guest houses in the Gigiri area. Please refer to the below list and make reservations at the hotel / guest house of your preference.

Recommended Hotels:

1. Windsor Golf and Country Hotel
2. The Tribe
3. Safari park Hotel (not so close to the Gigiri area)

Recommended Guest houses within Gigiri:

1. Tara Suites
2. Bedelle
3. Daisy Homes
4. Gigiri Homestead (54)
5. Casa Bella Suites
6. The 140
7. Shanema Homes
8. Dream Place

9. Kingdom Gardens
10. Villa Leone Hotels
11. Adies Garden Suites
12. Dolfran Guest House

Please kindly note that the transfer from the hotel to the course venue is responsibility of the participants.

Travel arrangements

Participants are responsible for their own travel and transfers.

You may plan your travel to arrive in time for the start of the course on Tuesday, 27 June 2017 at 9:00 am and may depart after the course closure on Thursday, 29 June 2017 around 5:30pm.

Visa

It is important that participants check passport and visa requirements for Kenya in due time before departure.

Holders of a valid UN Laissez-Passer (minimum 6 months) are not subject to the visa obligations.

Participants who do NOT hold a valid UN Laissez-Passer are requested to urgently check the requirements and time for visa application, if needed.

Kindly note that you might be able to obtain visa upon arrival at the airport, however, we would recommend that you apply for an e-visa online at <http://evisa.go.ke/evisa.html>.

Also, please note that Nationals of Senegal and Cameroon require referred visas unless they are holders of valid UNLPs. i.e. Kenya Visa must be applied about 3 (three) weeks before the travel date.

Nationals of Rwanda do not require visa to enter Kenya. However, it is advisable for the participants to have a copy of the invitation with them just in case it is required.

LOGISTICAL INFORMATION

Most countries need an e-visa to enter Kenya and some have further requirements. Please be sure to check the relevant visa requirements if you are traveling into Kenya for the workshop as soon as possible.

Transfers on arrival & departure

Hotels offer the service and may organize your transfer from and to the Jomo Kenyatta International Airport. Kindly request the pick up while confirming your room. Full travel details are required.

Local currency

The currency in Kenya is the Kenyan Shilling (KES)
The May 2017 UN Exchange rate is 1 USD = 103.4 KES

PRE-WORKSHOP PREPARATION

This worksheet is intended to help you prepare for the workshop by focusing on what you want to gain from the experience. **You can use it to reflect and collect your ideas, questions, and goals before the workshop, then bring it with you to help you track your learning over the four days.** You may also want to discuss these questions with a colleague or two before the workshop so that you can hone in on aspects of the workshop that are most useful to your organization.

1 Where are you starting?

What experience or skills related to Big Data are you starting with? What challenges are you facing? What questions do you have?

2 What do you hope the workshop will help you do?

List a few concrete ways that you hope the workshop will help you in your work.

3 What are ways you can get the most from the workshop experience?

Consider a few actions you could take to help you get the most from the workshop, such as connecting with speakers with expertise in your priority topics.

4 What can you contribute to the group?

The workshop will bring together a group of engaged professionals actively seeking to use Big Data for social good. What unique perspective, experience, and ideas can you share?

BIG DATA LITERACY PRIMER

What is data?

There is no agreed-upon definition of data. In general, data is an object, variable, or information that has the perceived capacity to be collected, stored, and identified. According to Oxford Dictionaries, data is “**facts and statistics collected together for reference or analysis.**” 1

STRUCTURED

- 1 Hierarchal structure
- 2 Least flexible
- 3 ~10% of data and decreasing
- 4 Each unit corresponds with a specific row and column, i.e. hierarchy. Follows ACID model: Atomicity, Consistency, Isolation, Durability

UNSTRUCTURED

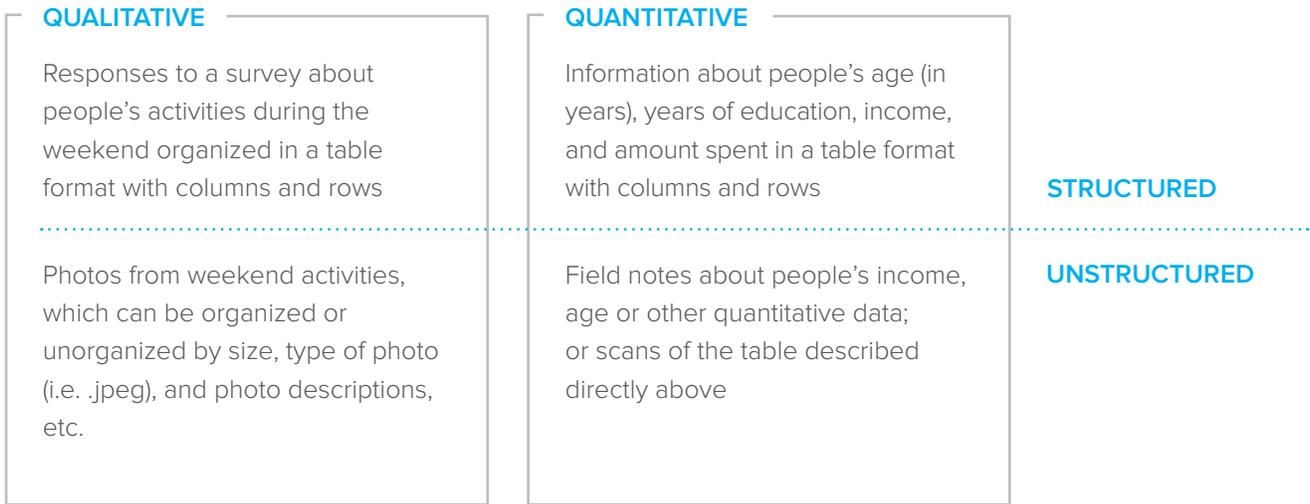
- 1 No set internal structure
- 2 Most flexible
- 3 ~90% of data and increasing
- 4 Each unit may have its own identifiable set of information and does not correspond to a particular hierarchy, such as film clips, pictures, and text documents

There are two main types of data: **structured** and **unstructured**. The former are created intentionally to answer a particular question asked prior to the data collection stage; as a result they are easy to search for, organize, and identify, and have a strict hierarchy. The hierarchy for a person’s favorite food might be: food, fruit, apple, red delicious. Each variable is clearly defined and labeled in a way that fits the structure’s taxonomy. Relational databases, popularized by IBM in the 1970s and 1980s, offered a significant improvement in the use of structured data in comparison to earlier hierarchical models.

Unstructured data are everything else. It can be photos, word documents, and other variables that do not need to follow a hierarchical method of identification. For example, someone can input data, such as an ‘apple’, without having to sequence it under the category of ‘fruits’ or know that there is a subcategory of ‘red delicious.’

Is unstructured data completely disorganized then? No. **Metadata** can be used to describe unstructured data. This can be **.jpeg** for example if it is used to describe a picture of an apple.

Over 90% of data is unstructured data, and it is growing exponentially in comparison to structured data because of the rapid creation of digital data, such as videos and tweets. As a World Bank report notes, “a 10-minute video of cats uploaded on YouTube may be quite heavy in terms of bytes but arguably contain fewer insights into the human experience than say Walt Whitman’s **Leaves of Grass.**” 2



The Big Data revolution is a result of this rapidly growing unstructured data. Much of this unstructured data is qualitative, however, the large majority of tools used to derive insights from data are quantitative in nature, such as statistics³. As greater techniques and tools are needed to analyze and make use of this data – i.e. actionable insights – there is a greater need to create quantitative, structured metadata surrounding these unstructured Big Data sets in order to employ these analytical tools.

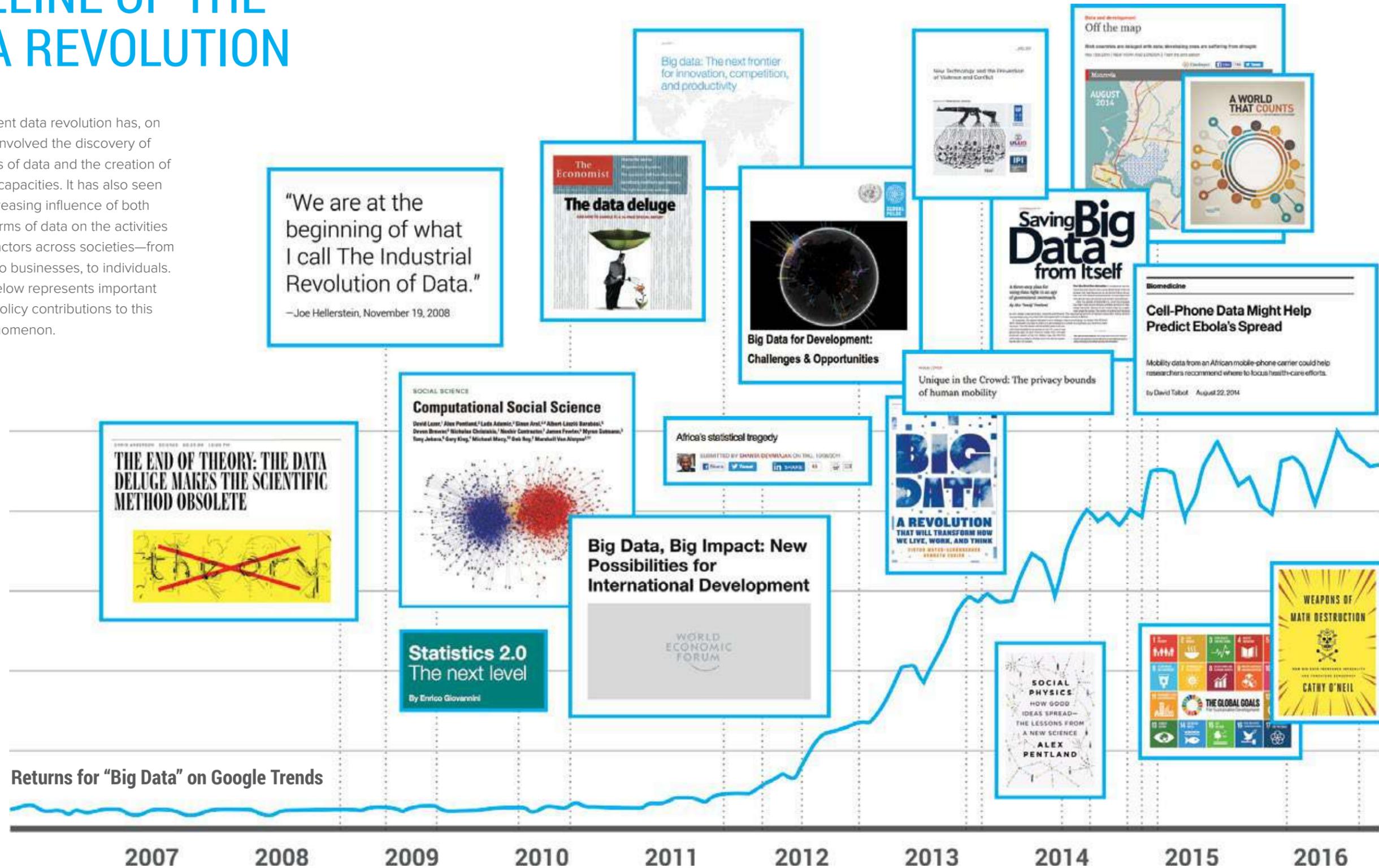
1 Oxford Dictionaries. Oxford Dictionaries Language Matters’ definition of “data”. Accessed September 2015. http://www.oxforddictionaries.com/us/definition/american_english/data

2 Hilbert, Martin. “How Big is Big Data?” Input to the World Bank 2016 World Development Report. June 2015: 2.

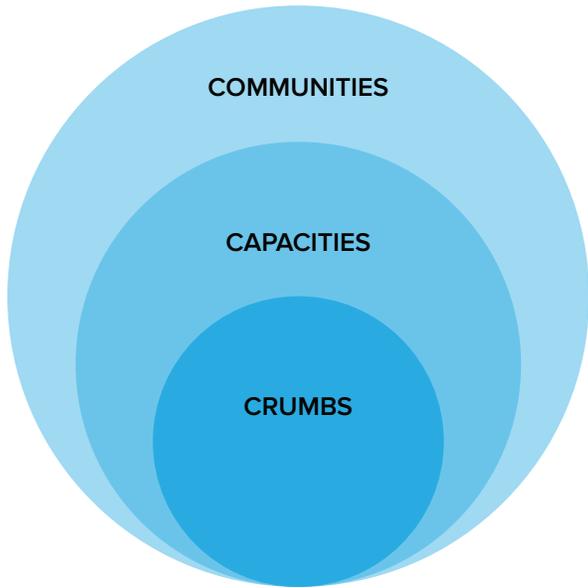
3 The American Statistical Association uses M. Davidian and T.A. Louis’ definition of “statistics” as “the science of learning from data, and of measuring, controlling, and communicating uncertainty; and it thereby provides the navigation essential for controlling the course of scientific and societal advances.” Science. 2012 Apr 6;336(6077):12. doi: 10.1126/science.1218685.

TIMELINE OF THE DATA REVOLUTION

The development data revolution has, on the one hand, involved the discovery of alternative uses of data and the creation of new analytical capacities. It has also seen the sharply increasing influence of both old and new forms of data on the activities and beliefs of actors across societies—from governments, to businesses, to individuals. The timeline below represents important research and policy contributions to this emerging phenomenon.

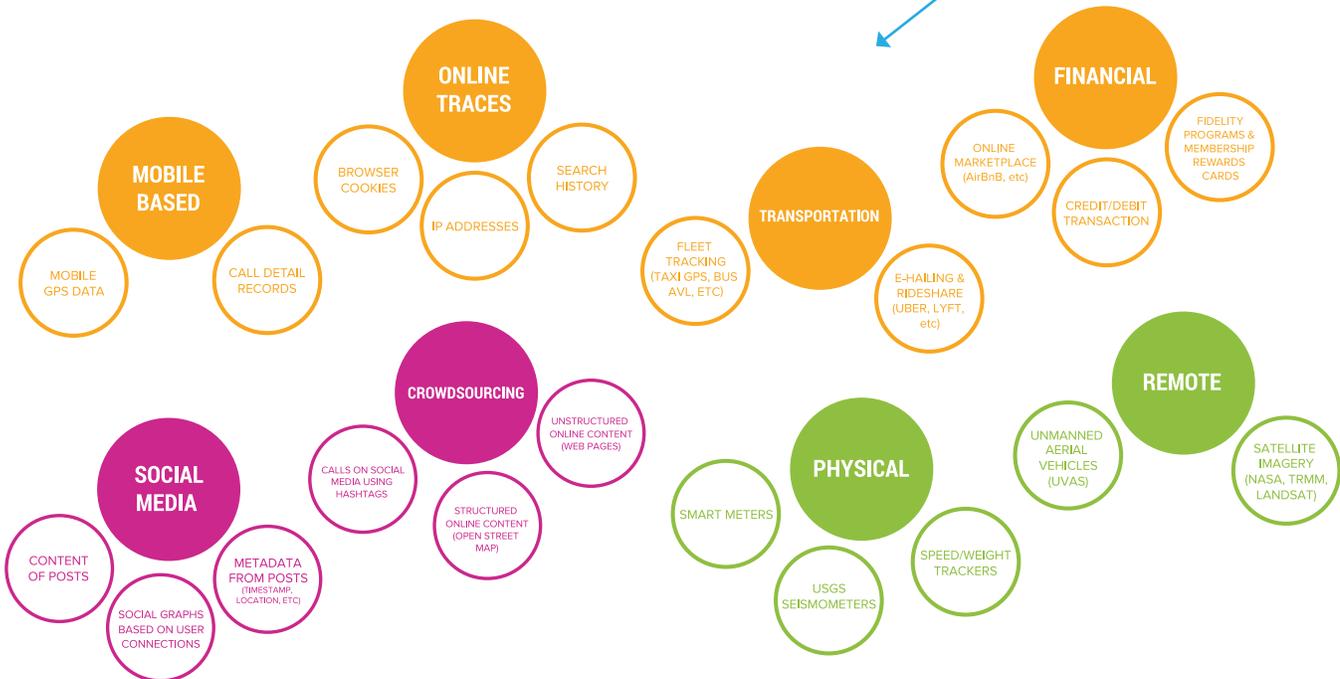
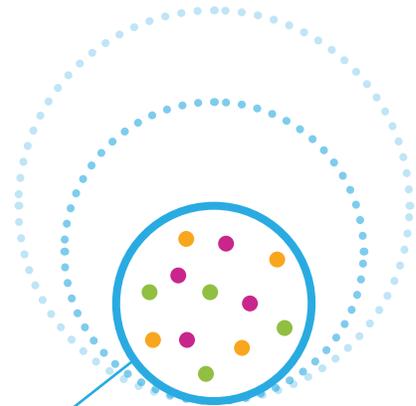


THE 3 C's OF BIG DATA

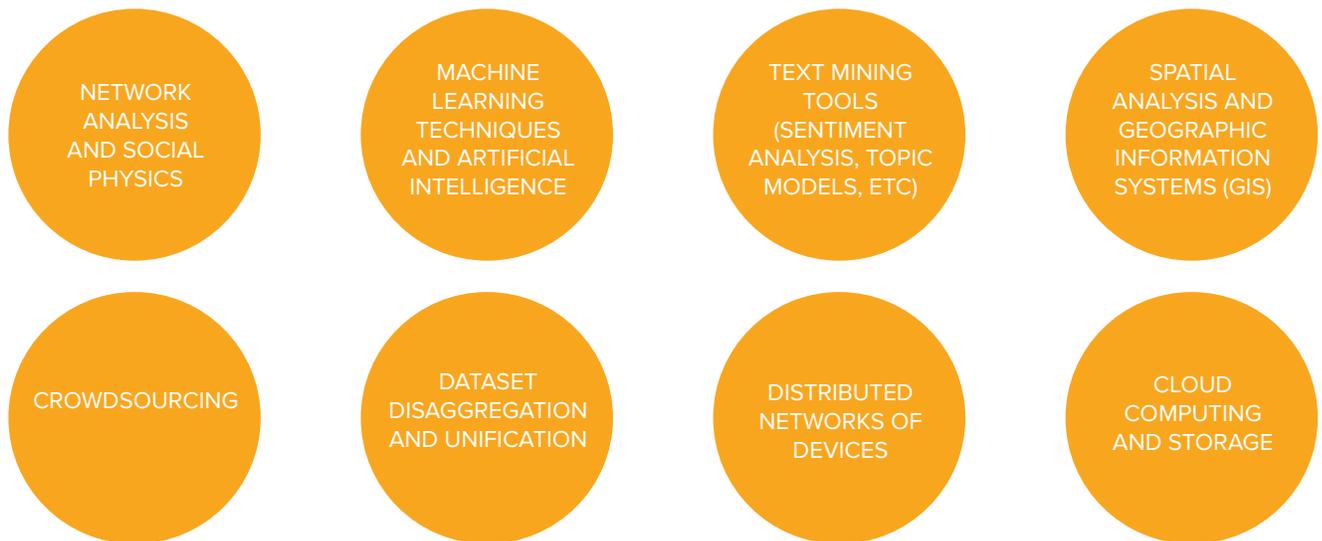
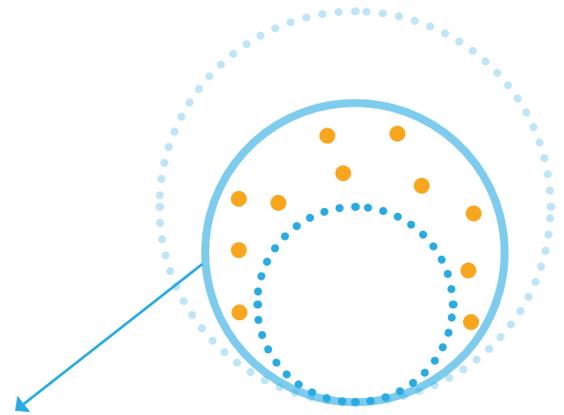


Originally framed as the “3 V’s” (volume, velocity and variety) in the early 2000s, we consider Big Data as an ecosystem of “3 C’s”: digital “crumbs” i.e. digital translations of human actions and interactions captured by digital devices; powerful capacities to collect, aggregate and analyze data; and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

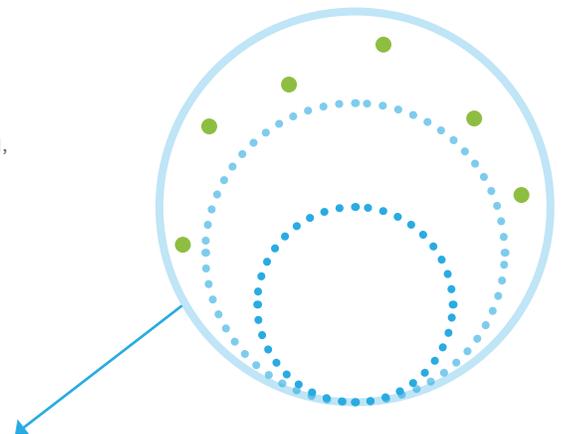
The **C of crumbs**—i.e. “digital bread crumbs” (a term coined by Alex “Sandy” Pentland), translations of human actions and interactions both passively emitted and captured by digital devices and other devices. At the center of our information societies is the production of massive amounts of data through 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, analyze and visualize. ‘Capacities’ act as a bridge between the crumbs and communities, helping to analyze both innovative, alternative uses of data as well as affecting how Big Data communities engage with and crowd around new knowledge and methodologies.



The **C of communities**—i.e. all stakeholders involved in generating, governing and using data, including data producers, end users, policymakers, civil society, experts, privacy advocates and civic hacker communities, as well as anyone represented in a dataset.





C²

CONTEXTS & CONCEPTS

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

Key readings

boyd, danah and Crawford, Kate, *Six Provocations for Big Data* (September 21, 2011) *A Decade in Internet Time: Symposium on the Dynamics of the Internet and Society*. Available online: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1926431 In this essay, the authors offer six provocations to spark conversations around issues of Big Data. As diverse groups argue about the potential benefits and costs of analyzing new data sources, left behind as digital traces and deposit data, significant questions emerge on this phenomenon and on its assumptions and biases. Will data analytics contribute to greater and better access to information by society, or will it trigger greater inequality? This paper was presented at Oxford Internet Institute's "A Decade in Internet Time: Symposium on the Dynamics of the Internet and Society."

Key learning points

- 1 Decode key terms and buzzwords in the Big Data and development landscape
- 2 Discuss Big Data within the broader political context of the post-2015 sustainable development framework and data for social good
- 3 Translate development problems into specific data objectives

Pentland, "Sandy" Alex, *Reinventing Society in the Wake of Big Data* (2012, August 30) Available online: https://www.edge.org/conversation/alex_sandy_pentland-reinventing-society-in-the-wake-of-big-data

This brief article features a conversation with Alex "Sandy" Pentland, Director of MIT Connection Science and the Human Dynamics Lab, at the MIT Media Lab. Starting with the power of Big Data about people's behavior, Alex "Sandy" Pentland explains further the consequence analysis of Big Data, what breadcrumbs tell us about people's choices, and the Big Data issues society should be considering. The article also provides a brief opinion on a data-driven society and issues concerning data ownership and control.

Letouzé, E. “Big Data for Development: Challenges and Opportunities”, (2012), UN Global Pulse. Available online: <http://www.unglobalpulse.org/sites/default/files/BigDataforDevelopment-UNGlobalPulseJune2012.pdf> This paper describes the potential for Big Data sources to allow decision makers to track development progress, improve social protection, and understand where existing policies and programmes require adjustment. The paper outlines key features of Big Data sources for development as digitally generated, passively produced, automatically

collected, geographically or temporally trackable, and continuously analyzed. With the promise of these new data sources come questions about the analytical value and thus policy relevance of this data—including concerns over the relevance of the data in developing country contexts, its representativeness, its reliability—as well as the overarching privacy issues of utilising personal data. This paper does not offer a grand theory of technology-driven social change in the Big Data era, but rather aims to delineate the main concerns and challenges raised by “Big Data for Development.”

Selected resources

VIDEOS

- 1 Pentland, A. “Sandy.” (2012). Reinventing Society in the Wake of Big Data: A Conversation with Alex “Sandy” Pentland. Edge.org. Available online: https://www.edge.org/conversation/alex_sandy_pentland-reinventing-society-in-the-wake-of-big-data
- 2 On Big Data, SDGs and Data Ecosystems: Emmanuel Letouzé, Director and Co-Founder Data-Pop Alliance. Available online: <https://www.youtube.com/watch?v=JYey1AJyLB8>
- 3 PARIS21: “Engineering a Development Data Revolution” Available online: <https://www.youtube.com/watch?v=-4LKaE83Cak>

ARTICLES AND PAPERS

- 1 World Economic Forum. (2012). Big Data, Big Impact: New Possibilities for International Development . World Economic Forum. Available online: <http://www.weforum.org/reports/big-data-big-impact-new-possibilities-international-development>
- 2 Independent Expert Advisory Group on a Data Revolution for Sustainable Development. (2014). A World That Counts: Mobilizing The Data Revolution for Sustainable Development (Data Revolution Group). United Nations Independent Expert Advisory Group on a Data Revolution for Sustainable

Development. Available online:

<http://www.undatarevolution.org/report/>

- 3 Letouzé, E. (2012). Big Data for Development: Challenges & Opportunities. UN Global Pulse. Available online: <http://www.unglobalpulse.org/sites/default/files/BigDataforDevelopment-UNGlobalPulseJune2012.pdf>
- 4 Mayer-Schönberger, V., & Cukier, K. (2013). Big Data: A Revolution That Will Transform How We Live, Work, and Think. Boston: Houghton Mifflin Harcourt
- 5 Pentland, A. “Sandy.” (2013). The Data-Driven Society. Scientific American, Available online: <http://connection.mit.edu/wp-content/uploads/sites/29/2015/01/Pentland-SciAmerican-Data-Driven-Society.pdf>
- 6 Pentland, A. “Sandy.” (2015). Social Physics: How Social Networks Can Make Us Smarter (Published with a new preface). New York, NY: Penguin Books.
- 7 The World Bank, World Bank Group, & Second Muse. (2014). Big Data in Action for Development. Available online: http://live.worldbank.org/sites/default/files/Big%20Data%20for%20Development%20Report_final%20version.pdf

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

Key readings

Bengtsson L, Lu X, Thorson A, Garfield R, von Schreeb J. Improved Response to Disasters and Outbreaks by Tracking Population Movements with Mobile Phone Network Data: A Post-Earthquake Geospatial Study in Haiti (August 2011) *PLoS Med* 8(8): e1001083. Available online: <http://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1001083>. This study uses SIM card data in Haiti to monitor population movements both before and after the 2010 earthquake and resulting cholera outbreak. It illustrates that, in regions with heavy mobile phone use, disaster-related population movements can potentially be quickly and accurately mapped using mobile phone data.

Key learning points

- 1 Understand existing methods and tools used to leverage Big Data
- 2 Assess data representativeness, biases and insights within Big Data-driven approaches and methods
- 3 Identify applicable tools by assessing the value add of Big Data for specific development problems

de Montjoye, Y.-A., Hidalgo, C.A., Verleysen, M. & Blondel, V.D. Unique in the Crowd: The privacy bounds of human mobility. *Nature* serp. 3, 1376; DOI:10.1038/srep01376 (2013). Available online: <https://www.ncbi.nlm.nih.gov/pubmed/23524645>.

The study on human mobility patterns emphasizes how human movement patterns are highly idiosyncratic and therefore, create unique traces that can be analyzed and identified based on databases that, in principle, contain anonymized data. Analyzing fifteen months of mobile phone data relating to roughly 1.5 million individuals, the authors of this study reveal that, when information is provided hourly about an individual via mobile phones, only four data points are needed to re-identify an individual.

Zagheni, Emilio, Weber, Ingmar, You are where you E-mail: Using E-mail Data to Estimate International Migration Rates (June, 2012) WebSci 2012. Available online: <http://dl.acm.org/citation.cfm?id=2380764>. The article suggests the use of e-mail data to complement existing migration data, resolving inconsistencies

arising from different definitions of migration, and providing new and rich information on mobility patterns and social networks of migrants. Additionally, the approach taken is to highlight the potential that the use of digital records have for developing countries, where the diffusion of Internet will be faster than the development of mature demographic registration systems.

Selected resources

VIDEOS

- 1 On the Future of Official Statistics in the Big Data Era: Nuria Oliver, Chief Data Scientist, Data-Pop Alliance, during a public panel discussion organised by the Overseas Development Institute, Data-Pop Alliance and the Royal Statistical Society. Available online: <https://www.youtube.com/watch?v=ucMznWQjLPM>
- 2 Big Data in Climate Research | Tomorrow Today. Available online: <https://www.youtube.com/watch?v=vbjAlGiHuk4>
- 3 Flowminder Foundation's work in response to the Nepal earthquake. Available online: <https://www.youtube.com/watch?v=hY9g0kPi8BA>

ARTICLES AND PAPERS

- 1 Baldrige, J. (2015). Machine Learning And Human Bias: An Uneasy Pair. Available online: <http://techcrunch.com/2015/08/02/machine-learning-and-human-bias-an-uneasy-pair/>
- 2 Blumenstock, J. E. (2012). Inferring Patterns of Internal Migration from Mobile Phone Call Records: Evidence from Rwanda. *Information Technology for Development*, 18 (2), 107–125. Available online: <http://doi.org/10.1080/02681102.2011.643209>
- 3 Gandomi, A., Haider, M., (2015). Beyond the Hype: Big Data Concepts, Methods, and Analytics. *International Journal of Information Management*. Volume 35, Issue 2. DOI: Available online: <http://doi.org/10.1016/j.ijinfomgt.2014.10.007>

- 4 Haklay, M., (2010). How Good is Volunteered Geographical Information? A comparative study of OpenStreetMap and Ordnance Survey datasets. *Environ Plann B Des* vol. 37 no. 4 682-703. DOI: Available online: <http://doi.org/10.1068/b35097>
- 5 Lazer, D., Kennedy, R., King, G., Vespignani, A., (2014). The Parable of Google Flu: Traps in Big Data Analysis. *Policy Forum*. Available online: <http://gking.harvard.edu/files/gking/files/0314policyforumff.pdf>
- 6 Pham, P., Vinck, P., Marchesi, B., Johnson, D., Dixon, P., and Sikkink, K., (2016) Evaluating Transitional Justice: The Role of Multi-Level Mixed Methods Datasets and the Colombia Reparation Program for War Victims, *Transitional Justice Review: Vol. 1: Iss. 4, Article 3*. Available online: <http://ir.lib.uwo.ca/tjreview/vol1/iss4/3>
- 7 Wallach, H. (2014). Big Data, Machine Learning, and the Social Sciences. Available online: <https://medium.com/@hannawallach/big-data-machine-learning-and-the-social-sciences-927a8e20460d>
- 8 Wesolowski, A., Eagle, N., Noor, A. M., Snow, R. W., & Buckee, C. O. (2013). The Impact of Biases in Mobile Phone Ownership on Estimates of Human Mobility. *Journal of The Royal Society Interface*, 10 (81), 20120986. Available online: <http://doi.org/10.1098/rsif.2012.0986>
- 9 Zagheni, E., Weber, I. (2015). Demographic Research with Non-representative Internet Data. *Emerald Insight*. Available online: http://www.zagheni.net/uploads/3/4/4/7/34477700/zagheni_weber2015.pdf



D*S

DESIGN & STRATEGY

Operationalizing Big Data as inclusive projects, partnerships and policies

Key readings

Bogomolov A, Lepri B, Staiano J, Letouzé E, Oliver N, Pianesi F, Pentland A (2015) Moves on the street: Classifying crime hotspots using aggregated anonymized data on people dynamics. Big Data 3:3, 148–158, Available online: <http://www.ncbi.nlm.nih.gov/pubmed/27442957>. In this article, the authors highlight the potential societal benefits derived from Big Data applications with a focus on citizen safety and crime prevention. The authors use demographic information along with human mobility characteristics as derived from anonymized and aggregated mobile network data and The authors explore the use of aggregated human behavioral data captured from the mobile network infrastructure, in combination with basic demographic information, to predict crime. The findings, built on and evaluated against real crime data from London, obtain

Key learning points

- 1 Identify individual and organizational objectives towards a Big Data strategy
- 2 Understand how to operationalize Big Data as projects, partnerships and policies
- 3 Recognize individual and organizational next steps towards Big Data applications

accuracy of almost 70% when classifying whether a specific area in the city will be a crime hotspot or not in the following month.

UN Data Revolution Group, A World that Counts: Mobilising the Data Revolution for Sustainable Development (November 2014). Available online: <http://www.undatarevolution.org/report/>. This report sets out major opportunities and risks presented by the data revolution for sustainable development. Prepared at the request of the United Nations Secretary-General, it calls for an urgent mobilization of the data revolution for all people, listing key recommendations to monitor progress, hold governments accountable, and foster sustainable development in the midst of this new phenomenon.

Wesolowski, A., Eagle, N., Tatem, A. J., Smith, D. L., Noor, A. M., Snow, R. W., & Buckee, C. O. (2012). Quantifying the impact of human mobility on malaria. *Science (New York, N.Y.)*, 338(6104), 267–270. Available online: <https://www.ncbi.nlm.nih.gov/pubmed/23066082>. Identifying the sources and sinks of imported malarial infections due to human travel and locating high-risk sites of parasite importation

could greatly improve malaria control programs. The authors used spatially explicit mobile phone data and malaria prevalence information from Kenya in order to identify the dynamics of human carriers that drive parasite importation between regions. The analysis identifies importation routes that contribute to malaria epidemiology on regional spatial scales.

Selected resources

VIDEOS

- 1 Orange's Data for Development Challenge - Senegal. Available online: <https://www.youtube.com/watch?v=LmndoB1iXg4>
- 2 Introduction to the Global Partnership for Sustainable Development Data. Available online: <https://www.youtube.com/watch?v=Og7GB0jqm2I>
- 3 WIRED 2013 - Nuria Oliver: what Big Data and the Mexican pandemic taught us. Available online: https://www.youtube.com/watch?v=H5_FeuuS-zs&spfreload=5

ARTICLES AND PAPERS

- 1 Ballivian, A., & Hoffman, W. (2015). Public-Private Partnerships for Data. Available online: http://data.worldbank.org/sites/default/files/issue-paper-financing-the-data-revolution-ppps_0.pdf
- 2 Ceola, S., F. Laio, and A. Montanari (2014). Satellite Nighttime Lights Revealing Increased Human Exposure to Floods Worldwide, *Geophys. Res. Lett.*, 41, 7184–7190, doi: Available online: <http://doi.org/10.1002/2014GL061859>
- 3 de Montjoye, Y. A., Smoreda, Z., Trinquart, R., Ziemlicki, C., & Blondel, V. D. (2014). D4D-Senegal: the second mobile phone data for development challenge. arXiv preprint arXiv:1407.4885.

- 4 Fournier-Tombs, E. UNDP, Human Development Report (2015). Visualizing Human Development Data: Achieving a Fine Balance Between Science and Art. Available online: <http://hdr.undp.org/en/content/visualizing-human-development-data-achieving-fine-balance-between-science-and-art>
- 5 Harvard School of Public Health. (2012). Using Cell Phone Data to Curb the Spread of Malaria. Available online: <https://www.hsph.harvard.edu/news/press-releases/cell-phone-data-malaria/>
- 6 How the citizen data scientist will democratize big data". (2016). Available online: <http://www.forbes.com/sites/bernardmarr/2016/04/01/how-the-citizen-data-scientist-will-democratize-big-data/#7157fca45573>
- 7 Innova Challenge Big Data; an example of open innovation. (2014). BBVAOpen4U. Retrieved 31 October 2016, Available online: <https://bbvaopen4u.com/en/actualidad/innova-challenge-big-data-example-open-innovation>
- 8 Tatem, A. J. et al (2014). Integrating Rapid Risk Mapping and Mobile Phone Call Record Data for Strategic Malaria Elimination Planning. *Malaria Journal* Available online: <http://malariajournal.biomedcentral.com/articles/10.1186/1475-2875-13-5>



ETHICS & ENGAGEMENT

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

Key readings

Letouzé E, Vinck P. “The Law, Politics and Ethics of Cell Phone Data Analytics.” *Data-Pop Alliance White Paper Series*. Data-Pop Alliance, World Bank Group, Harvard Humanitarian Initiative, MIT Media Lab and Overseas Development Institute. April 2015. Available online: <http://datapopalliance.org/item/white-paper-the-law-politics-and-ethics-of-cell-phone-data-analytics/>. This paper examines Call Detail Records (CDRs), a record that includes among other data, the starting time of the call (or message), its duration, the originating and receiving phone numbers, and the benefits and risks involved in their sharing and use. CDRs are essentially used for billing, monitoring voice and data usage, and understanding and targeting customers based on their cell phone consumption patterns, but has recently been recognized for the insight they provide into human behavior, movements, and social interactions. As the proposed use of CDR expands well beyond their original purpose, certain ethical and legal questions need to be addressed. The paper summarizes current legal

Key learning points

- 1 Identify models for prioritizing inclusivity, transparency and accountability in data public-private-people partnerships
- 2 Articulate and assess ethical, privacy and legal implications of Big Data applications
- 3 Understand key principles for effective data communication and story-telling

frameworks before exploring structural socio-political parameters and incentives structuring the sharing of CDRs, proposing guiding ethical principles, and discussing operational options and requirements.

“Beyond Data Literacy: Reinventing Community Engagement and Empowerment in the Age of Data.” *Data-Pop Alliance White Paper Series*. Data-Pop Alliance (Harvard Humanitarian Initiative, MIT Media Lab and Overseas Development Institute) and Internews. September 2015. Available online: <http://datapopalliance.org/item/beyond-data-literacy-reinventing-community-engagement-and-empowerment-in-the-age-of-data/>. The term ‘data literacy’ has gradually emerged as a mainstream term and potential buzzword of the ‘Data Revolution’ discussions, as experts, policymakers and advocates began considering what it would take to enable citizens to make better use of the vast amount of data available to them. The authors of this paper begin

by discussing ‘data literacy’ as an emerging concept within a much longer historical narrative of literacy promotion, and define data literacy as “the desire and ability to constructively engage in society through and about data.” In utilizing a definition of data literacy that builds on the elements of current sub-categories of literacy and expands beyond particular media—and their technocrats—the authors describe four key pillars that form its foundation: data education, data visualizations, data modelling, and data participation. The paper further considers what it would mean to be “literate in the age of data” and denote four core pillars in literacy promotion.

Nissenbaum, Helen. *Privacy in Context: Technology, Policy, and the Integrity of Social Life* (June 2016) Stanford: Stanford UP, 2009. 304. ISBN: 9780804772891 In this book, Helen Nissenbaum argues on the nature of people’s complaints when it comes to their privacy: they are not necessarily against the act of sharing information itself, but against the inappropriate, improper sharing of it. She continues by stressing the importance of having governing norms on distribution and protection in distinct social contexts, and how contemporary information systems should raise an alarm when they weaken the fabric of social life.

Selected resources

VIDEOS

- 1 Big Data: Key to the future or the end of privacy? | Data-Pop Alliance and Vodafone Institute’s digitising Europe Initiative, Brussels Debate featuring Linnet Taylor and Kenneth Cukier. Available online: <https://www.youtube.com/watch?v=F2pPXDn-46E>
- 2 WIRED 2015 | How Big Data will Help us Hold Governments Accountable by Alex Sandy Pentland. Available online: <https://www.youtube.com/watch?v=VZTLJOiBGs8>
- 3 Big but personal data: Yves-Alexandre de Montjoye, Research Affiliate, Data-Pop Alliance at TEDxLouvainLaNeuve. Available online: <https://www.youtube.com/watch?v=DPqreIYe0UU>

ARTICLES AND PAPERS

- 1 Albright, A., Levine, S., (2015). One Size Does Not Fit All: The Shortcomings of the Mainstream Data Scientist Working for Social Good. Available online: <https://thelittledataset.files.wordpress.com/2015/09/albright-levine-one-size-does-not-fit-all.pdf>
- 2 Barocas, S., & Selbst, A. D. (2015). Big Data’s Disparate Impact (SSRN Scholarly Paper No. ID 2477899). Rochester, NY: Social Science Research Network. Available online: <http://papers.ssrn.com/abstract=2477899>
- 3 Crawford, K., & Schultz, J. (2013). Big Data and Due Process: Toward a Framework to Redress Predictive Privacy Harms (SSRN Scholarly Paper No. ID 2325784). Rochester, NY: Social Science Research Network. Available online: <http://lawdigitalcommons.bc.edu/cgi/viewcontent.cgi?article=3351&context=bclr>

- 4 D’Ignazio, C., Bhargava, R. (n.d) Approaches to Building Big Data Literacy. Available online: http://www.kanarinka.com/wp-content/uploads/2015/07/Big_Data_Literacy.pdf
- 5 Greenwood, D. et al. (2014). The New Deal on Data: A Framework for Institutional Controls. ISBN: 9781107590205. DOI: Available online: <http://dx.doi.org/10.1017/CBO9781107590205.012>
- 6 Homeland Security, Science and Technology. (2012). The Menlo Report: Ethical Principles Guiding Information and Communication Technology Research. Available online: https://www.caida.org/publications/papers/2012/menlo_report_actual_formatted/
- 7 O’Neil, C. (2016). Weapons of Mass Destruction: How Big Data Increases Inequality and Threatens Democracy. ISBN 0553418815
- 8 Stikeleather, J. (2013). How to Tell a Story with Data. Harvard Business Review. Available online: <https://hbr.org/2013/04/how-to-tell-a-story-with-data/>
- 9 Pasquale, F. (2015). The Black Box Society. The Secret Algorithms That Control Money and Information. ISBN 9780674368279
- 10 Saurwein, F., Just N., Latzer, M. (2015). Governance of Algorithms: Options and Limitations. Available online: http://www.mediachange.ch/media/pdf/publications/GovernanceOfAlgorithms_.pdf
- 11 Taylor, L. (2016). Mobile Phone Data Sharing in Emergencies - Consent, Care and Control. Available online: <https://linnetaylor.wordpress.com/2016/03/15/mobile-data-sharing-in-emergencies-consent-care-and-control/>
- 12 United Nations Human Rights. Office of the High Commissioner. (n.d.) The Right of Privacy in the Digital age. Available online: <http://www.ohchr.org/EN/Issues/DigitalAge/Pages/DigitalAgeIndex.aspx>

GLOSSARY OF TERMS

Algorithms

In mathematics and computer science, an algorithm is a series of predefined instructions or rules written in a programming language designed to tell a computer how to sequentially solve a recurrent problem through calculations and data processing.

Big Data

An umbrella term signifying one or more of three trends: the growing volume of digital data generated as a by-product of the use of digital devices by people on a daily basis; new technologies, tools and methods available to analyze large datasets not originally designed for analysis; and the intention to extract from these data and tools insights that can be used for policymaking. In this primer, Big Data is distinguished from big data, where Big Data refers to a new ecosystem made up of its 3 C's for Big Data Crumbs, Capacities and Community.

Data

An object, variable, or piece of information that has the perceived capacity to be collected, stored, and identified. Data come largely in two forms: structured data – which are highly organized, identifiable, and often hierarchical – and unstructured – which do not need to follow a hierarchical method of identification and are often not as readily amenable to automated analysis. (See explanation on pp. 10-11).

Data literacy

The desire and ability to engage constructively in society through and with data. Data literacy incorporates elements and principles from various sub-kinds of literacy (such as media, statistical, scientific computational, information and digital literacies) and encompasses the set of tools and skill that individuals need in order to understand, interpret, and manage data-driven decisions and arguments, particularly in the context of the 'Data Revolution'.

Data modeling

Using existing datasets to infer current conditions or predict future outcomes. The process involves resolving complex relationships among datasets in order to understand what the data means and how the elements relate together.

Data revolution

A common term in development discourse since the High-Level Panel of Eminent Persons on the Post-2015 Development Agenda called for a 'data revolution' to "strengthen data and statistics for accountability and decision-making purposes". It refers to a larger phenomenon than Big Data or the 'social data revolution' — defined as the shift in human communication patterns towards greater personal information sharing, and the implications of this.

Data science

A field that focuses on solving real-world problems using large amounts of data by combining skills from often distinct areas of expertise: maths, computer science (for example, hacking and coding), statistics, social science and even storytelling or art.

Data sovereignty

The right of individuals to set the rules of what happens to data generated by and about themselves.

Demographically identifiable information (DII)

All forms of data which, in isolation or through linking, can lead to the identification, classification, and tracking of demographic groups. This includes personal identifiable information (PII), online data, geographic and geospatial data, environmental data, survey data, and census data.

Digital divide

The differential access and ability to use information and communications technologies between individuals, communities and countries — and the resulting socioeconomic and political inequalities. The skills and tools required to absorb and analyse the growing amounts of data produced by such technologies may lead to a ‘new digital divide.’

Exhaust data

Data that are passively emitted from cell phones, sensors, social media and other platforms as digital translations of human actions and interactions.

Group privacy

The concept of privacy applied to groups. This concept is separate and different from that of individual privacy, since there can be situations when the privacy of a group is at risk even as its members’ individual privacy is protected. The discourse surrounding whether and how to define group privacy is twofold: should group privacy be a legally enforceable right; and if so, (how) can it be protected in domestic or international legal frameworks?

Open data

Data that is easily accessible, machine-readable, accessible for free or at negligible cost, and with minimal limitations on its use, transformation, and distribution.

Privacy

A notoriously difficult to define concept, here viewed as a facet of human dignity: one’s right to have a measure of knowledge and control over what information is made public about oneself.

Small data

‘Hard,’ structured data that can be easily quantified and organized (in columns and rows for instance) for systematic analysis, and that cannot be edited by their emitters. Examples include information about people’s age, years of education, and income in a table format with columns and rows.

(Statistical) Machine learning

A subset of data science, falling at the intersection of traditional statistics and machine learning. Machine learning refers to the construction and study of computer algorithms — step-by-step procedures used for calculations and classification — that can ‘learn’ when exposed to new data. The addition of “statistical” reflects the emphasis on statistical analysis and methodology, which is the main approach to modern machine learning.

Web-scraping

A computer software technique for automating the extraction of information from websites.

an illustrated introduction to

Predicting socioeconomic levels through cell-phone data

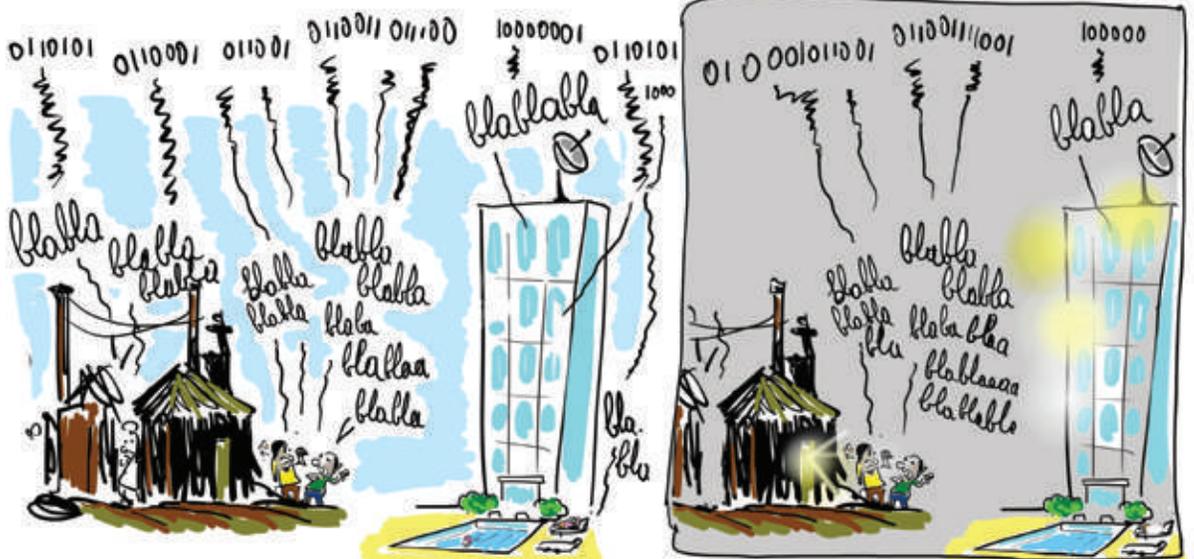
Question:



step ①



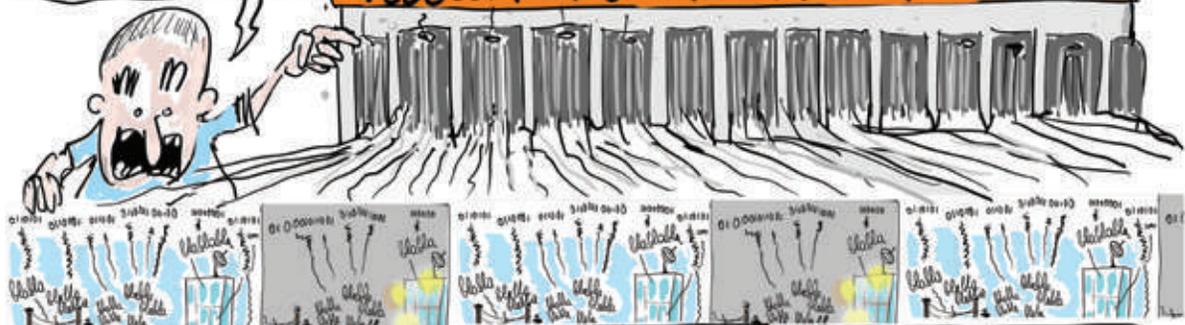
step ②



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

CALLER ID	CALLER LOCATION	RECIPIENT ID	RECIPIENT LOCATION	CALL TIME	CALL DURATION
X36872	2°24'22"	A8C492	3°38'49"	2014.04.01	01.12.27
9748Y	35°49'56"	TC73646	31°12'22"	ET 17 22	

TELECOM OPERATOR DATA CENTER



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



