What is data justice for (South) Africa

and who gets to decide?

Linnet Taylor
TILT, Tilburg University
Netherlands
5.3.2019
‘When it was proclaimed that the Library contained all books, the first impression was one of extravagant happiness... There was no personal or world problem whose eloquent solution did not exist in some hexagon.’

Jorge Luis Borges (1941)
The Library of Babel
The machine readable society

By 2030, profiling will be based on data gathered over decades

By 2130, profiling will be based on data gathered over generations
Digitisation and new data technologies are making lower-income-country populations visible in new ways.
Data for development and crisis response, 2010

- Epidemiology
  - Flowminder

- Crisis response-
  - Telenor mobile data for Bangladesh floods

- ‘Nowcasting’ economic shocks through airtime purchases
  - UN Global Pulse

- Dengue forecasting
  - Telenor/Harvard partnership in Pakistan
United Nations (2014)
Citizens or consumers?

Nigeria: biometric ID card + Mastercard + integrated data (driving licences, voter registration, health insurance, taxes and pensions)
The commercial population database: India’s Aadhaar

- World’s largest biometric database: 1.22bn entries by 2018

- Adopted in order to reduce corruption in welfare system, but run by nonstate actors

- ‘Non-compulsory’ but functionally necessary

- Function creep occurring toward:
  - Governing (proof of legal residence)
  - Shaping access to opportunity (scholarships, exams, schools)
  - Marketing (direct monetisation of data)
Data volume = power

Volume vs. quality: 165 acknowledgement slips (by 2014) with photos of non-humans
Commercially provided human rights?

The ID2020 initiative: blockchain vendors claim they are providing ‘basic rights’

1.1 BILLION PEOPLE LIVE WITHOUT AN OFFICIALLY RECOGNIZED IDENTITY

THIS LACK OF RECOGNIZED IDENTIFICATION DEPRIVES THEM OF PROTECTION, ACCESS TO SERVICES, AND BASIC RIGHTS. ID2020 IS A PUBLIC-PRIVATE PARTNERSHIP DEDICATED TO SOLVING THE CHALLENGES OF IDENTITY FOR THESE PEOPLE THROUGH TECHNOLOGY.

https://id2020.org
Using mobile network data (specifically the communication patterns as well as the history of airtime credit purchases) from Orange in Côte d'Ivoire, researchers estimated the relative income of individuals, and the diversity and inequality of income. They then collectively used these measures to understand socioeconomic segregation at a fine-grained level for Côte d'Ivoire. The above figure shows the poor areas (in blue) in relation to the areas of high economic activity (yellow to red areas).

(Gutierrez et al. 2013)
Are social media/mobile traces ‘development data’?

‘Big Data are not about society, but about users and markets. They are therefore inherently biased in that they do not track people who fall outside the particular markets or activities being tracked’

(Shearmuir 2015)
Questions raised by global datafication

1. Individual vs. collective targeting: how do we know what data we care about?

2. A duty of visibility: do we have a right to invisibility?

3. Who is in charge?: government and private sector merge
1. individual vs. collective targeting

‘we are only Big Brother to the masses, we don’t identify anyone personally’

(Albert Seubers, Eindhoven ‘Stratumseind’ living lab, 2015)
De-identified data is safe data

‘The principles of data protection should therefore not apply to anonymous information... This Regulation does not therefore concern the processing of such anonymous information, including for statistical or research purposes.’

(GDPR, 2016)
Problem: we don’t always know what kind of data we care about

my data

Provided:
- Name
- Political party membership
- Mobile #

Collected:
- Phone metadata
- GPS location history
- Satellite data

Data about me

Data about people like me

Inferred:
- Consumer preferences
- Political affiliation
- Risk quotient
Group privacy (Taylor, Floridi & vd Sloot 2017)

- Traditional view: a group is a collection of privacies
  - A corporation (firm, university, state)
  - A collective (political party, social network)

- Data analytic perspective: a group is fluid, based on a chosen predicate
  - People on a bus
  - Students at the University of Cape Town

- Data analytics capture types based on proxies, and are used predictively
Non-identifiable data tell a story

• Reflect our identity-building process, as well as freedom and autonomy (Floridi 2013)

• Provide ways to predict our behaviour, and that of others

• Data which are anonymous (or which never conveyed individual identities) can be constitutive. They can tell us about the lives of groups, networks, villages, cities, states
2. Do we have a duty to be visible?
‘Not using data is the moral equivalent of burning books’

(Kenneth Cukier, Data editor, The Economist, 2016)
‘We see big data as a real public good.’

(Robert Kirkpatrick, UN Global Pulse, IGF panel 2013)
Can privacy be traded for other important rights?

‘Privacy is your right. So is access to food, water, humanitarian response. The challenge is that we see a lot of regulatory frameworks which don’t have the right litmus test.’

(Robert Kirkpatrick, UN Global Pulse, 2014)
[capabilities and human rights] are ‘trumps’ in the sense that they have a very strong priority over the pursuit of welfare generally: if any one of them is not fulfilled, the nation is not even minimally just.

(Nussbaum, 2011)

NB: Human rights cannot be claimed against firms or multilaterals
3. Government and market data are merging
Is exposure to the market part of the social contract?

‘Mobile phones are a storehouse of personal data and reflect upon individual preferences, lifestyle and choices. The conflation of biometric information with SIM cards poses grave threats to individual privacy, liberty and autonomy...The mere existence of a legitimate state aim will not justify the disproportionate means...’

(Supreme Court of India, Aadhaar judgement, September 26 2018)
What is data?

• An asset

• Labour

• Social relations

• A building block of our identity

... how we govern it depends on how we frame it
Options for governing the intersection of state and commercial data

1. Pretend it’s not happening
2. Focus on trust and ethics
3. Treat data as an ecosystem
4. Take a data justice perspective
Proposition 1: pretend it’s not happening
Proposition 2: focus on trust and ethics

Do They Trust You with Their Data?
Percentages of consumers who said that each category of organization was “trustworthy” or “completely trustworthy” when it came to making sure that personal data was never misused.

- **87%**  
  Primary care doctors

- **85%**  
  Payment or credit card companies

- **80%**  
  E-commerce firms

- **77%**  
  Consumer electronics firms

- **76%**  
  Insurance companies

- **76%**  
  Banks

- **73%**  
  Telecom carriers

- **70%**  
  Technology firms

- **68%**  
  Internet giants (such as Google)

- **66%**  
  Government entities

RESPONSIBLE DATA SCIENCE

5 ethics principles big data analysts must follow

We trust big data and its processing far too much, according to Altmetric analysts. Get tips on incorporating ethics into your analytics projects.

By Michael Kassner   January 2, 2017, 6:00 AM PST
From rules to trust: embrace the grey area

- ‘Trust’ and ‘ethics’ imply a particular story about the data market:
  - Citizens can identify and engage with the technologies involved
  - Citizens can choose, limit, change those technologies
  - Firms can prioritise ethics in choices about technology and data
  - Citizens can verify how these choices are being made

- ‘data ethics’ may be:
  - An indicator of missing rules
  - A process of ad hoc standard setting
  - A statement about self-regulation
Proposition 3: treat data as ecosystem

• Data as ‘system resource’, an ecosystem of people, platforms and profiles, with distributed ownership (Purtova 2015):
   e.g. ‘Personal data ecosystem’ (Hamlin & Hodder, 2011; World Economic Forum, 2014)

• Draws attention to the flaws of the ‘responsible individual’ paradigm

Social justice-based approach to data governance:

- Empirical approach to understanding what people need from data technologies, seeking common principles across cultures

- Situates the right to be seen and the potential disciplinary role of data systems within the social contract

- Incorporates different rules/rights paradigms
A justice-based approach might...

- Prioritise human needs over market efficiency

- Address governance as a dynamic problem (e.g. risk modelling, expert review processes, institutional evolution) rather than static (guidelines and codes)

- Account for different definitions and interests (data as relations, data as a component of the self)

- Be based on legitimacy rather than consent
Seeking common principles

- Visibility
- Engagement with technology
- Nondiscrimination

DATA JUSTICE
References


