Open Data & Digital Identity: Lessons for Aadhaar

Amba Kak
Mozilla Foundation (amba@mozillafoundation.org)

Smriti Parsheera
National Institute of Public Finance & Policy (smriti.parsheera@nipfp.org.in)

Nanjing, China
27-29 November 2017
The case for more open data

• What is “open data”? : openness is on a spectrum

• Open data as a public good - accountability, research, business decisions

• Moving beyond government data BUT - Are there incentives to “open”?  
  – Government: right to information laws, open data policies  
  – Private sector: voluntary frameworks, MiData case study

• Why is Aadhaar - the largest biometric identification system in the world - an appropriate case study?
An introduction to Aadhaar

• What is Aadhaar?

• Enrolment of residents:
  – 1.18 billion people enrolled (Nov)
  – 85% of Indian population in 2017*

• Applications of Aadhaar:
  – Distribution of welfare services/ payments
  – Financial sector (the “JAM” trinity)
  – Telecom sector
  – Emerging uses: health, education.

Source: *State of Aadhaar report
How it works: Data flows for authentication & eKYC

eKYC*

Authentication**

* https://aadhaarapi.com/
**State of Aadhaar Report
Open data & privacy

• Privacy safeguards already hardcoded in law:
  – Provisions in Information Technology Act
  – Exception of privacy in RTI law
  – Aadhaar law contains confidentiality provisions

• The global debate on open data, privacy, anonymity

• Privacy debate as it is unfolding in India:
  – Aug 2017 Supreme Court decision on privacy
  – Committee on data protection law

Where should the balance lie?
Sources of open data in Aadhaar

- UIDAI: enrolment, authentication, failure rates
- User agencies: public and private

**Authentication Trend**

- Number of authentications:
  - Dec 2016: 0
  - Jan 2017: 300 M
  - Feb 2017: 600 M
  - Mar 2017: 900 M
  - Apr 2017: 1,200 M
  - May 2017: 1,500 M
  - Jun 2017: 2,800 M
  - Jul 2017: 3,000 M
  - Aug 2017: 5,358 M
  - Sep 2017: 7,004 M
  - Oct 2017: 10,152 M
  - Nov 2017: 13,000 M

- Cumulative authentications:
  - Dec 2016: 0
  - Jan 2017: 300 M
  - Feb 2017: 900 M
  - Mar 2017: 1,500 M
  - Apr 2017: 3,000 M
  - May 2017: 5,358 M
  - Jun 2017: 10,152 M
  - Jul 2017: 13,000 M

**eKYC Trend**

- Number of eKYC transactions:
  - Dec 2016: 0
  - Jan 2017: 200 M
  - Feb 2017: 300 M
  - Mar 2017: 400 M
  - Apr 2017: 500 M
  - May 2017: 600 M
  - Jun 2017: 800 M
  - Jul 2017: 1,000 M
  - Aug 2017: 1,200 M
  - Sep 2017: 1,400 M
  - Oct 2017: 1,600 M
  - Nov 2017: 1,800 M

- Cumulative eKYC transactions:
  - Dec 2016: 0
  - Jan 2017: 200 M
  - Feb 2017: 500 M
  - Mar 2017: 900 M
  - Apr 2017: 1,300 M
  - May 2017: 1,700 M
  - Jun 2017: 2,100 M
  - Jul 2017: 2,500 M
  - Aug 2017: 2,900 M
  - Sep 2017: 3,300 M
  - Oct 2017: 3,700 M
  - Nov 2017: 4,000 M
Some potential applications

• Telecom subscriber data
  – gender divide in technology?
  – distribution across districts?

• Hospital registration data
  – what is the incidence of disease?
  – age and gender of patients?

• Transgender community
  – extent of exclusion?
Key principles and implementation

• Key principles:
  – Personally identifiable information must not be disclosed
  – Begin with aggregated, anonymised data
  – Develop further standards with time

• Who decides what is to be shared and how?
  – Sector regulators, UIDAI, data collectors themselves?

• Our proposal (For Aadhaar and beyond)
  – An independent mechanism
  – Multi-stakeholder approach
  – Oversight of implementation processes
Proposed framework

• Role of the open data committee
  – Recommend contractual commitments - *will vary from sector to sector*
  – Develop principles and standards for data sharing
  – Follow open, transparent processes
  – Monitor compliance by agencies
  – Communications strategy for disseminating information

• UIDAI makes decisions, informed by open data committee
Objective

Open data for better research, enhanced accountability and design improvements in the Aadhaar ecosystem.

Thank you