On health data architecture design

Subhashis Banerjee

Computer Science and Engineering
IIT Delhi

Law Economics Policy Conference, Delhi
November 27, 2018
Why do we need it?

- Electronic Health Records and individual health.
- Public health monitoring.
- Socio-economic studies.
- Epidemiology.
- Research.

Caveat: Data cannot be substitute for fundamentals - PHCs, doctors, . . .

Concerns: Privacy; potential imbalance between private and public.
Stunting in (North) India

Source: NFHS-4
Outline

▶ Objectives of a health data system design
▶ Operational considerations
▶ Design considerations: alternatives
Objectives: Health records (what should be recorded?)

- Birth record and certificate.
- Immunisation records.
- PHC records, all medical episodes, prescriptions and doctor’s opinions.
- Tests, imaging, radiology and pathology reports.
- Hospital case records, discharge certificates.
- Lifestyle indicators (dietary habits, smoking, drinking, activity patterns), chronic conditions.
- Optionally record history of self-medication (quackery included), home measurements of BP, sugar, etc., Garmin, FitBit and other wearables.
- Genetic data?
- Death record and certificate.
Objectives: Health records and access

- **Federated** data collection and management of health records.
- **Individual centric** architecture. Individual is the data controller.
- Ensure that no access to health records is possible without consent.
- Exceptions? **Authorised accesses**? Mandatory/selective disclosures? Emergency overrides? Limited access to parent/sibling PHRs?
- **Selectively** grant read/write access to health professionals, hospitals, test and imaging centres, insurance.
- All accesses to be **logged** in a non-repudiable and immutable manner.
- **No duplicated data** at hospitals, PHCs? Restrict post-treatment access?
Objectives: Analytics

- Regular operational surveillance (anomalies and alerts)
  - Epidemic and endemic conditions like dengue, malaria, TB, cholera, typhoid, …
  - Malnutrition, vitamin or other micro-nutrient deficiencies in populations and regions.

- Epidemiological studies
  - Purpose specific analytics.

- Research and non-profit studies
  - May require aggregated digests and anonymised longitudinal data.

- Commercial research
  - Anonymised and aggregated data.
  - Data economics? Consent? Payments?
Objectives: Privacy and security

- **Access control**
  - Only *programmatic access* through secure APIs.
  - Only legitimate and authenticated access, enabled by *consent and authorisation*.
  - No unauthorised linking with other data and personal identifiers.
  - Non-repudiable and immutable logs of all accesses. Access control also for the logs.

- **Purpose limitation**
  - Ensure that *no access violates purpose of consent or authorisation*.
  - Accessing programs to *provide proofs/guarantees*.
  - *ex-ante* rather than *ex-post*.

- **Regulatory framework**
Operational considerations: things to watch out for

- **UHID** derived from a **national digital identity**.
- **Digital literacy? Network access** in remote areas? **Authentication, authorisation** and consent methods?
- Local caching of data?
- **Inter-operability**. Data portability or where-is?
- APIs and use cases.
- **Standardisation** and **inter-operability** of software and Apps at PHCs, hospitals and clinics, imaging and test centres, pathologists, radiologists.
- **A comprehensive law** harmonious with digital identity, data protection and IT Acts. No money bill please!
Design considerations: Blockchains?

- **Permissioned Blockchains** to maintain non-repudiable logs of all data generation and data access.

- PHR data compartmentalised and **encrypted with** a hierarchy of **personal keys**. User in control of data.

- Each consensus participant maintains all data; either in **monolithic databases**, or in **decentralised**, distributed, fault tolerant, peer-to-peer file systems such as the IPFS.

- **Consent** and **authorisation** architecture based on **smart contracts**.

Design considerations: Blockchains?

▶ Advantages:
  ▶ Transparency, correctness, non-repudiation, immutable.
  ▶ Basic framework well tested and standard (except scalability, of course).
  ▶ Can support federated generation of information.
  ▶ Multiple central authorities (miners).
  ▶ Distributed protocol (but not really decentralised in terms of storage and computations).
  ▶ Can support APIs.
  ▶ Smart contracts natural for consent and authorisations architecture.
Design considerations: Blockchains?

- **Disadvantages:**
  - State capacity?
  - PoW or BFT may require *excessive redundant computation*? Power plants?
  - Still require **strong regulatory framework** for access control (prevent bypass of access through smart contracts) and **purpose limitation**. Centralised DPA? Replicated at each consensus participant?
  - Support **access for analytics** through smart contracts? Private keys? Centralisation? Devil lies in details?
Design considerations: Monolithic?

▶ **Advantages:**
  - Easier, from a state-capacity point of view.
  - Can be made secured, fault tolerant.
  - Regulated access control and purpose limitation easier to implement?
  - Non-repudiable and immutable through fault tolerance and regulated access control?

▶ **Disadvantages:**
  - Transparency.
  - Convincing people.
Design considerations: Others

- Interface design for individuals, PHCs, ... 
- Digital literacy? Interface design for consent.
- Methods for **access control** and **purpose limitation**.
- Limits of anonymization of medical data with guarantees against re-identification attacks?
- Key management. Reset? Hierarchy of master keys (Merkle tree based?) will imply centralisation.
- Connectivity? Caching design? Lazy commits?
- Above all, **whitpapers** and **public consultations**.