Proposed National Personal Data Protection Plan

P Rajagopal Tampi

Introduction

The collection, storage, control and use of personal data by Big Tech companies is fast becoming a central concern for the world.

It is an essential responsibility of the Government to protect personal data of citizens assured by the Constitution of India. The coming of AI threatens the lives, livelihoods, possessions and rights of citizens due to misuse of personal data and modelling individuals using AI resorted to by Big Tech Companies. The government would be protecting the rights of individual citizens and upholding constitutional guarantees by ensuring the protection of personal data of its citizens.

The common practice today of Big Tech Companies is to transfer personal data from mobiles and other personal devices without the complete knowledge, understanding and consent of the individual. The Terms and Conditions of use is couched in voluminous legalese which the lay person fails to comprehend and is meant to obfuscate through complicating the matters involved.

Personal data transfer to cloud-based servers enables Big Tech companies to train their personal AI agents and facilitates AI modelling of citizens and their behaviour patterns to earn even more profits. Having run out of learning data, Big Tech companies are focussing on the extracting data from the daily lives and behaviours of ordinary people. This is an illegal practice as it violates individual human rights and constitutional guarantees given by nations.

The swearing of President Trump and the “MAGA” fever, has introduced an age where information held by American companies about Indian citizens can be withheld on the order of the United States government[1]. This puts every Indian citizen in great danger. Therefore, it is of critical importance that India put in place a personal data protection infrastructure, which encompasses technology, policy, and processes to ensure that even if the US government mandates such illegal access to personal data of Indian citizens, it would not be able to achieve its purpose.

Definitions

API: Application Programming Interface.

Core Personal Data (CPD): Core Personal Data consists of all types of identification data of the individual. It also consists of the persons behavioural attributes, preferences, dislikes, psychometric profile, psychological attributes medical data and other information which can be used to model the individual using an AI model. Core personal data comprises data listed in table 1, very high risk category.

Controller of Citizen’s Personal Data (CCPD): This government organization will hold the registration of all vendor and partner organizations. It will be a testing and auditing centre for personal data protection of citizens. Statutory reports will be sent to it by vendors.

Public Personal Identifier (PPI): The public identity number allocated to a citizen of a country by the Government for personal data protection. As its name implies it is public information. Thus number will be used to anonymise personal data transfers for enhancing personal security.

Mapping Personal Identifier (MPI): Mapping Person Identifier (MPI) issued by the Government to each citizen is strictly confidential and stored only in the government MPI database. It is not known

even to the individual. It is used to construct the pointer to the Adhaar ID of the citizen when combined with the PPI.

UIDAI: Unique Identification Authority of India.

Personal Data Watermarking (PDW): technology to mark personal data elements to facilitate identification, categorization, tracking and auditing.

Risk classification of personal data

The levels of risk posed by personal data can be classified into four main categories. The category "very high risk" encompasses personal information which can be used for impersonation, may result in losing one's identity and for AI modelling of the individual. The second category is "high risk", which poses a threat to the individual's livelihood, possessions and work opportunities. The third category of risk is "medium risk" which includes obtaining an individual's personal information through indirect means such as through one's social interaction friends and relatives. The fourth category is low risk which includes peripheral information like leisure travel hobbies and sports. More detailed information on these risk categories is included in Table 1 below.

Risk Category	Personal Data falling into category
Very High (Core Personal Data)	Name, Address, Date of Birth, biometrics, facial images, voice signature, Passport number, Location, Psychometric and Psychological profiles, behaviour, likes, dislikes, preferences, value systems, medical and test reports, significant life experiences etc.
High	Data on Finances and Investments, debts, property, assets, liabilities, work experience, CV, employers feedback
Medium	Friends and relatives, Social interactions and affiliations,
Low	Leisure travel, hobbies and non-professional sports etc.

Table1: Personal Data Risk level categorisation table.

Personal Data Watermarking (PDW)

Meity should mandate implementation of Personal Data Watermarking (PDW) in all digital applications using unmasked personal data. This technique should facilitate identification of personal data, its category, creator details, modification details, tracking it life-cycle and permit auditing by the Controller of Citizen's Personal Data (CCPD).

Personal Data Security Technology and Governance Infrastructure

Public Person Identifier

To ensure the protection of individuals, their personal identification data and core personal data, the introduction of a Public Person Identifier (PPI) number for every citizen is recommended. The PPI is public data and it is entered into the form for opening an account with any vendor or service provider instead of the person's name. The other information which is required to be supplied by the customer is her email ID. With this PPI-based system, there is no need to enter the name, address, contact number or other details since these details will be obtained by the vendor in anonymized form from the UIDAI authority by providing the PPI and the e-mail ID of the customer.

The PPI is specifically generated and designed to identify the person using an external government controlled secure identity establishment service.

This PPI can be used for three purposes, namely legal identification when needed, recovery of anonymised KYC information, customer authentication, customer authorisation to transfer personal data and approving the purpose and extent of the individual's personal data to be transferred.

Whenever personal data is being transferred out of a personal device, if data is required for identification or authentication, it will not be sent with the person's name or any other identifier but with the unique Public Person Identifier (PPI).

When the personal device/system is programmatically (without manual intervention* or manual command**) transferring personal data out of the device for any purpose whatsoever, in case of the need to identify the person involved, the data will include a Public Person Identifier (PPI) in the place of (high risk) personal identifiers like name/PAN number etc. No other data including device identifiers like IMEI number, telephone number, Mac ID which can indirectly facilitate personal identification should be transmitted during non-manual data transfers/ transactions (i.e. purely device/system driven transfers/transactions). This will ensure that personal data exchange is always adequately anonymised. The data transferred will be anonymized data supplied by the CCPD as explained below.

The PPI will also facilitate identity recovery through the approved government identification process if required.

Obtaining anonymized KYC information

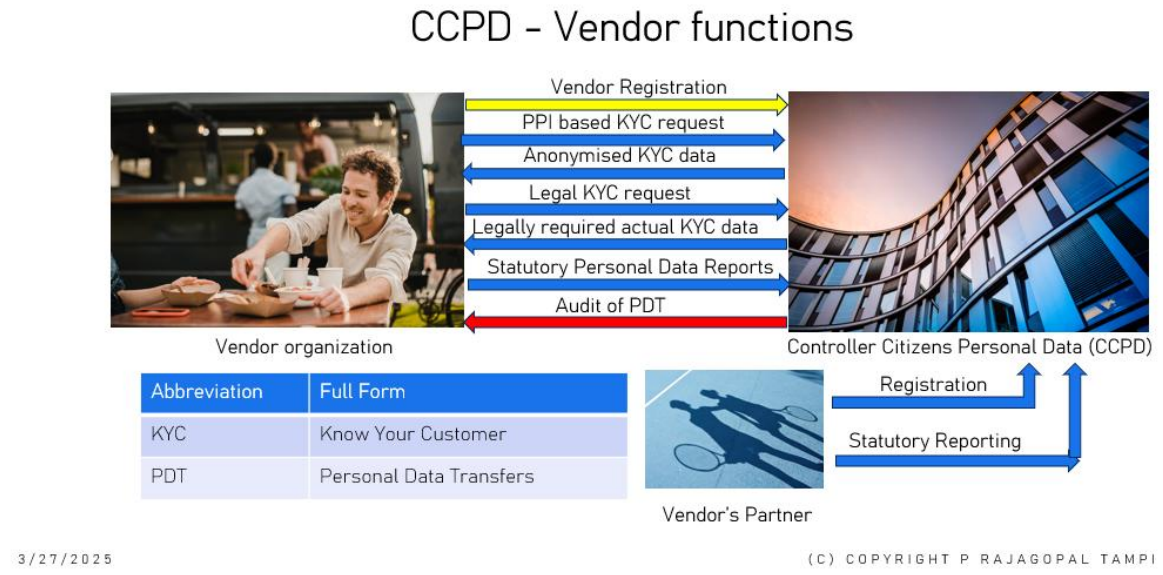

Fig 1: Interaction between CCPD and Vendors

The UIDAI will be the central organization to provide the infrastructure for obtaining anonymised KYC information to requesting vendors using PPI and email id as the inputs. The Public Person Identifier (PPI) can be shared openly with all agencies by the individual. The PPI has another associated part called the Matching Person Identifier (MPI) which is strictly known only by the government and stored in the government MPI database within UIDAI. It is not known even to the individual. The combination of PPI and MPI will point to the Adhaar number of the individual. The identity verification process will be carried out by decoding PPI >MPI>Pointer (PPI+MPI) to Adhaar address> Adhaar No. This Aadhaar discovery process starting with PPI is explained in figure to below.

It is critical that the PPI and the Adhaar number are never mentioned together in any government or other document. Else the security provided by this system becomes weakened greatly.

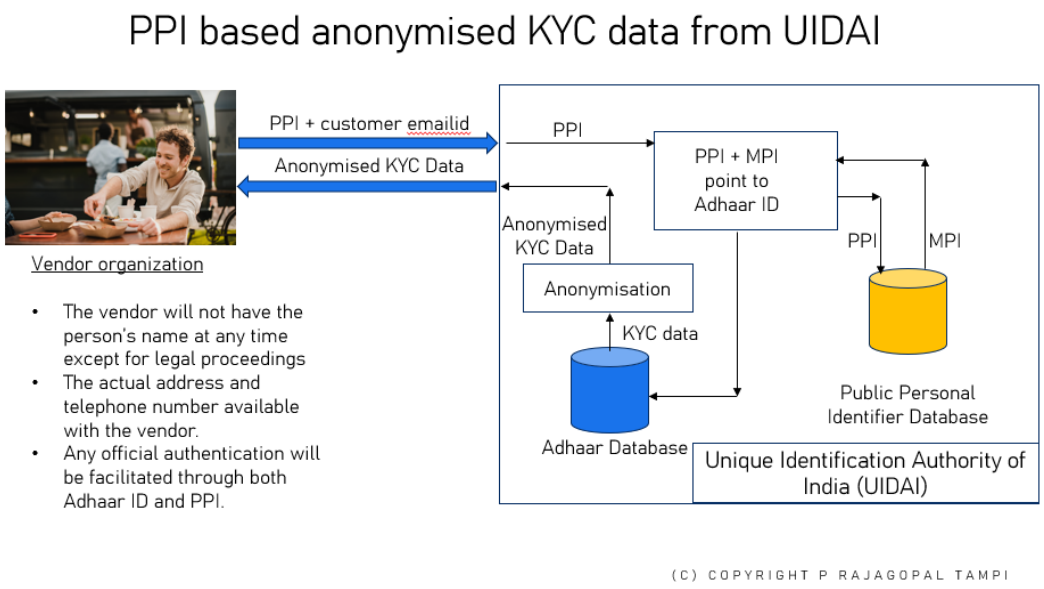

Fig 2: UIDAI extracting the KYC based on PPI supplied by the vendor

In case Aadhaar based person identification is legitimately required in some cases eg: legal proceedings, the current Aaadhar based identification system can be used. The aadhaar card identification process will involve the UIDAI system sending an otp to the individual concerned informing him of the agency which is requesting for identification and the purpose of identification.

Government Support Infrastructure

In India, the PPI and MPI can be generated by the Unique Identification Authority of India (UIDAI). The matching of the PPI with the aadhaar number will be carried out in memory of the government system and message regarding confirmation or otherwise will be sent to the requesting system or person. This discovery of Adhaar using the two-parts (PPI and MPI) will be designed so that PPI and MPI will reside in one database and aadhaar IDs will reside on a separate database as illustrated in Fig 2 above.

PPI and MPI will be generated by the UIDAI organisation. MPI will not be visible to citizens since it being used internally by the UIDAI for mapping PPI to Adhaar. PPI will be provided to all citizens in the same way as the Adhaar ID has been done.

Controller of Citizen's Personal Data (CCPD): This new proposed government organization will be set up to hold the registration of all vendor and partner organizations. It will be a testing and auditing centre for personal data protection of citizens. It will be the recipient of statutory reporting on personal data transfers made by vendors. This organization will be responsible for personal data protection audits of vendors and compiling reports to be made to the Executive.

Vendor Personal Data Processing

Every product vendor who requires to remove/export personal data from an individual's personal device is required to set up its own PDP technology infrastructure. This infrastructure will provide OTP and (2FA) authenticator mechanism which will authenticate and obtain the consent of the individual whose personal data is being transferred out of the personal device. The approval is required for each transfer. Each personal data transfer will be logged with details such as the purpose of data removal, to whom the data is being handed over to along with details of consent of the individual, data components transferred, and time validity of the data transferred etc. The CCPD is

authorised to audit these logs and report violations to appropriate government authority for necessary action against such vendors.

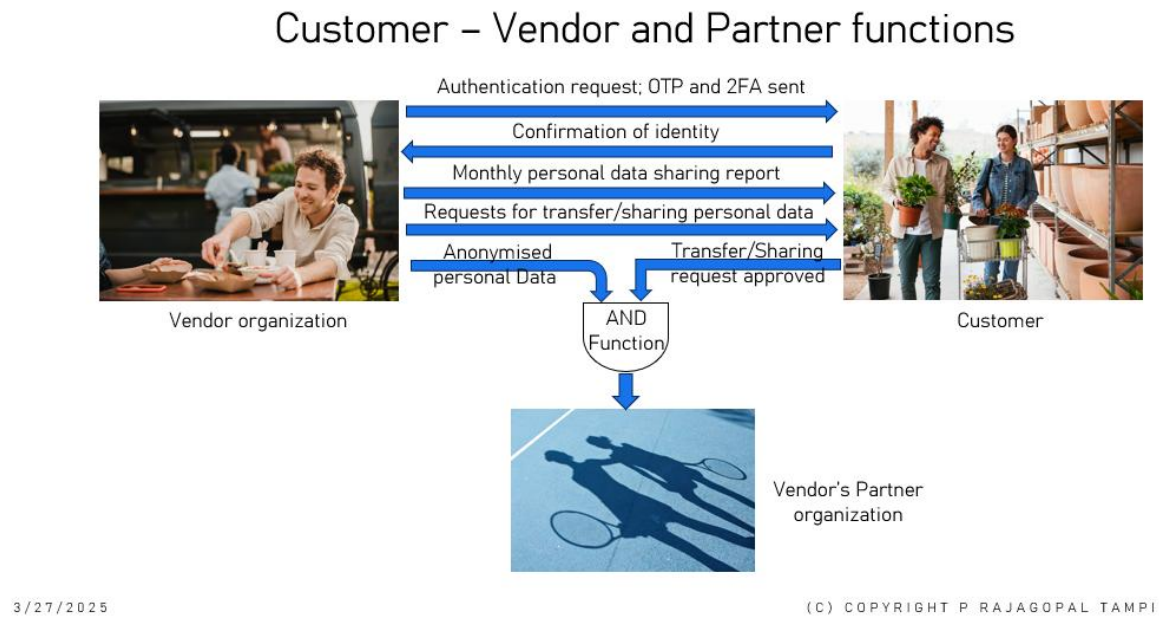

Fig 3: Process for transferring personal data to their partners by vendors

Core Personal Data – Formats, storing and transferring rules

1. Names will be stored in the anonymized format in which the KYC data has been supplied by CCPD. Dates of Birth will be supplied in month and year only without the date. Address and telephone numbers will be actual data. On request from the vendor, the CCPD will transfer only the minimum relevant KYC information required depending on the industry to which the vendor belongs.
2. Currently existing and new Email ids generated will be anonymized in a suitable manner for public purposes by the email service providers. The service providers may have mechanisms to decode the original email id behind the anonymised one for legal purposes including government use.
3. The recording of facial recognition, voice signatures and other Very High-risk category personal data of customers by the vendors must only take place after due consent of the individual in real time for every case through the authentication authorization process.
4. CPD should be stored by apps in encrypted format in the database. No app will provide access to another app's CPD internally or directly from the personal device. Any personal data transfer between apps must take place through a path which is external to the personal device i.e. the CPD must be transferred out of the device and read in again by the requesting app using the APIs of the personal device).
5. For CPD sharing between apps installed on the personal device, the vendor permitting such an operation is required to log the details of the transfer. Such logging will contain at the minimum PPI, partner name to whom transferred, customer consent log, purpose of transfer, data components transferred, how long the data will remain active etc.
6. Real time location data will be transferred only with the consent and approval of the individual concerned and for an approved duration only. Real time location sharing by vendor's will be documented and reported on in a separate category by the vendor to the CCPD.

7. Vendors will not encrypt or transform authorised personal data like address and telephone number into any other form and transfer them outside the device to obfuscate activities. Such action should be penalised heavily.
8. It must be possible for the individual at any time to instruct the vendor to delete or reset all bio-metric and other login features and other core personal data. The vendor shall comply with all requests within a published SLA.
9. Application Terms and conditions of Vendors: the complex and huge terms and conditions which users of software systems need to agree to today must be simplified by obtaining user consent for personal data transfers to external devices for every case in real time.
10. Reports (of personal data transfers) should be available as a menu item of the product/service accessible to Controller of Citizen's Personal Data roles and to device owners with following details:
    a. CPD components accessed and transferred, purpose with details.
    b. PPI based authentication and data transfer details.

Exceptions to PPI based KYC

Banks, Insurance companies, telecom companies, pension funds, stocks purchases and property investments, government agency registrations and transactions will be exempted from PPI based KYC. In these cases, the exact identity data needs to be disclosed. There are some other areas where personal data is concentrated in the private sector. These include human resource management (product and service) companies and contact list management products.

The personal data security regulations in these exempt areas needs to be tightened to prevent disclosure of personal data. The core approach should be to split the CPD data into separate parts with the aim of not being able to complete the CPD profile of the individual easily thus increasing the level of security.

Other applications which process personal data like "journal apps" should accept and log the permissions and the consent of users. Personal Data saved on the servers must be encrypted and not readable from the database access.

Progressive future personal data protection steps

All personal data will be tagged with risk categories. The tagging must be possible at data structure level meaning that the data byte/component itself can be parsed to identify its risk level in real time. The risk level is then mapped to government personal data laws which will decide its treatment by the device with respect to security aspects. These laws should be built into the device as an embedded (even using ASIC chip level) mapping within the operating systems and APIs transferring personal data to/from the device. Audit logging of all personal data export transactions should be available with full details including API used and API functions accessed for such data transfers.

References:

1. https://www.wired.com/story/trump-era-digital-expat/

Notes: * When a person fills out a form for opening a website account, or editing his particulars, this process is called manual intervention.

**When an AI agent is collecting and transferring information, the agent may have the generic permission of the data owner to collect information, but the agent cannot transfer this personal information out of the device without the owner's explicit per transaction real-time approval. This

approval will be logged and reported upon by the vendor.

The author is a pathfinder for AI. He is an Author, Investor, Entrepreneur, CEO and Veteran from the Indian Navy. He has worked in software from 1982, being responsible for many of the Indian Navy's simulators during the 1980's and 1990's. As a Captain in the Indian Navy, he was the Indian Defence Adviser in Nigeria and Ghana. He is an MTech (Computer Science) from IIT Bombay.

Author: P Rajagopal Tampi

Book: Applied Human Centric AI

Website: https://aipathfinder.org
