Data Management DATA9911: 2021-21 (Assignment 1)

Harsh Vardhan Rai

D20123653

D20123653@mytudublin.ie

TU059 – Data Science

Introduction

Under the Central Government of India, a contact tracing application was developed by the NIC to track the spread of COVID19 cases across the country. However, there were many controversies related to the implementation of this application and the data policies from the civil rights advocates, activists, hackers society, and even the political enthusiast. This debate was escalated at such a level from a political perspective that it was even called "a surveillance system with a private administrator" by the opposition parties **(Dewan, 2020).** Not getting deep into this political propaganda and personal views, by ground research I found that there were/are many voids regarding the ethical and legal issues during every release.

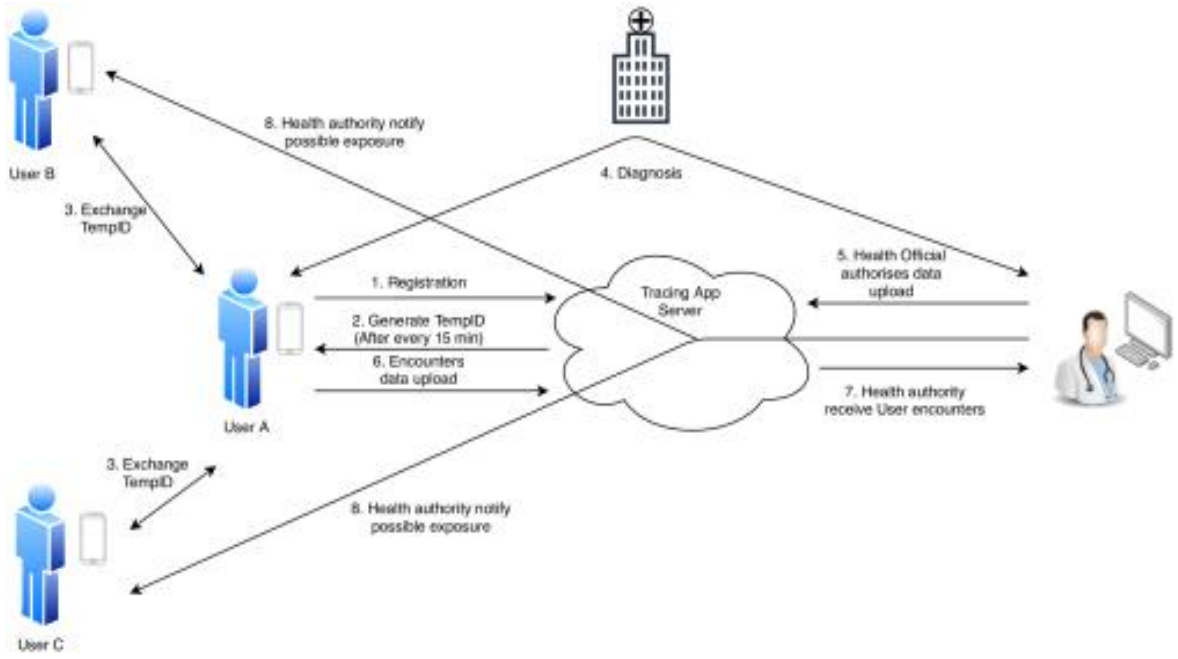
- There were proofs by an ethical hacker who goes by the name Elliot Alderson that there was an activity named WebView in the backend which is primarily responsible for viewing privacy policy but in the proof uploaded on Twitter, it can open even the internal files of the application and it can open any URL without any host validation **(Alderson, 2020)**.
- The rooted devices can't access the application but after some scripting in Frida after decompiling the application, we can easily bypass the certification **(Alderson, 2020).**
- According to the privacy policy and the front end of the application, we can limit the radius of individuals affected by COVID to only 5 options but in the backend, these parameters can be easily changed and any attacker can easily have access to any patient all over the country **(Alderson, 2020).**
- Privacy breach and no legislative sanction towards the usage of the application while making the application mandatory in all organizations **(Dewan, 2020)**.

Data Collection and Data Processing

- If we are talking about the data cycle of the application through which an individual gets advised to get tested or quarantine, we should look at key points through which it is processed which are:
    - What is the information which is collected?

- What technology is being used by the application for contact tracing?
- Where is the data stored?
- For how many days is the data kept in the system or storage?
- Who may have access to the data which is collected by the application?

- There are basic three types of system architecture which the tracing application proposes are *Centralized*, *Decentralized,* and *Hybrid*. Aarogya Setu here uses the ***centralized*** approach for data collection and processing.



Source: (Ahmed et al., 2020)

- The technology used here is Bluetooth and GPS. In this case, Bluetooth detects other devices (if the device is in the range) and exchanges a digital signature with the time stamp and the location coordinates. This data however remains in the backend. If that person or you test positive within the 14 days of interaction, the application will then provide you or that person with suitable inputs and medical interventions based on the range of that infected person.

- The data is stored in a central repository from the databases to analyze or manipulate it and there are several criteria through which the data is stored.
    - **30days**: If the information is not synced to the government servers. However, it will be there on the phone.
    - **45days**: Some information of the individuals might be there on the server who were not tested positive to implement various health responses.
    - **60days**: Some information of the person who was tested positive can be there on the server from the day he recovered.

Storage	Registration Phase	Operation Phase	(If tested positive)
On Server	Name, Phone number, Sex, Profession, Travel History(Last 30 days), Age	DID (Device ID)	Details related to the positive cases and the close contacts
On Device	NA	DID (Device ID), Phone model, Timestamp, Bluetooth, GPS, DIDs contact	NA

Table describing the data pipeline

- The data which is collected by the application can be used transparently by the Health Ministry (Health Departments of State/Union Territory/Local), National & State Disaster Management Authorities, Public Health Institutions (State/Union Territory/Local). Some anonymized data can also be used for research under the Universities that are registered in India.

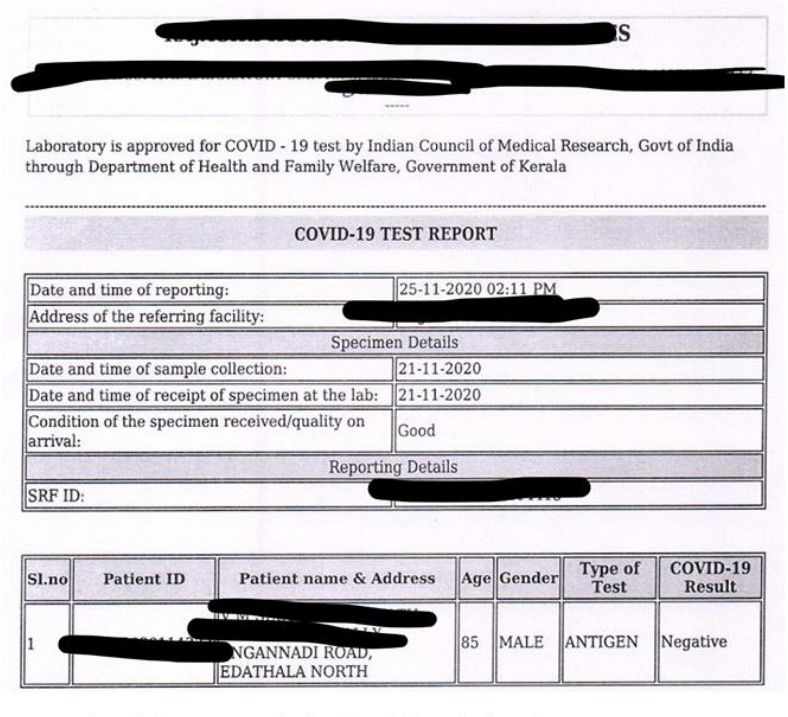
Data Governance, Privacy, and Ethics

Many questions were raised regarding the approach towards data storage, safeguarding individual privacy as the application was collecting too much personal

information, guidelines regarding the implementation, and lawful foundation for the data protection under the Personal Data Protection Act, 2019.

- The legal and policy domain of the application was questioned for its ToS (Terms of Service) stating that the data which has to be deleted after a particular number of weeks was not referring to all the data which was on the server. They stated that if the data is “anonymized” it can be further used for medical and administrative interventions **(Newsclick Report, 2020)**. But this anonymous data can also be questioned on the behalf as:
    - There are no particular standards defined for anonymizing the user data under the ToS.
    - No consent was taken from the user regarding their anonymized data.

- The EHR(Electronic Health Records) policies and protocols during an emergency/pandemic situation lack some standards. It can be supported by the fact that tens of thousands of user details were leaked on Government of Delhi domains in early January 2021. In the early stages of COVID19, the Karnataka government, on purpose, released the addresses of patients who tested positive to support contact-tracing **(Kurian, 2021b)**.



[redacted]

Laboratory is approved for COVID - 19 test by Indian Council of Medical Research, Govt of India through Department of Health and Family Welfare, Government of Kerala

COVID-19 TEST REPORT

Date and time of reporting:	25-11-2020 02:11 PM
Address of the referring facility:	[redacted]
Specimen Details	
Date and time of sample collection:	21-11-2020
Date and time of receipt of specimen at the lab:	21-11-2020
Condition of the specimen received/quality on arrival:	Good
Reporting Details	
SRF ID:	[redacted]

Sl.no	Patient ID	Patient name & Address	Age	Gender	Type of Test	COVID-19 Result
1	[redacted]	[redacted] NGANNADI ROAD, EDATHALA NORTH	85	MALE	ANTIGEN	Negative

Source: (Kurian, 2021a)

- Under the (Reasonable Security Practices and Procedures and Sensitive Personal Data or Information), Rules 2011, the government authorities are impertinent towards any privacy and protection framework.

- Data collection needs to be minimal and appropriate to maintain the issues related to user's privacy. In contrast, the application asks the user to enter the mandatory details such as Name, Phone number, Age, Sex, Profession, Countries visited in the last 30 days and the Location while in the backend it already has the MAC address (device through which we are logged in), GPS (Latitude and Longitude), Device ID (Static), Distance between the devices and the signal strength, Bluetooth model and name, and Time of synchronization with the contact device through the GPS and Bluetooth technology. The amount of data that is being collected here in contrast to the other applications in other countries is excessive and questionable, making it a trust deficit.

- According to the latest draft of PDPB, 2019 which was introduced in Parliament of India, it grants indemnity to GoI (Government of India) during various situations of pandemic or emergencies which can breach the right to privacy.

- While the user registers with the application, the data entered is linked to a Digital ID (DID). While locating the proximity, these DIDs are exchanged but the problem with these DIDs is that they are static and it becomes very easy to de-anonymize the user identity as it implements only a single encryption layer.

- The proposal of this application with an e-pass feature to access the public transport services of metro trains or movements during lockdown makes it mandatory to be installed thus eroding an individual's legitimacy as promised in the Constitution. Also in the application, a section has been added for the donations to the PM-CARES fund, which seems to have no limitations or legislative oversight to what this application stands for.

First Principle Ethical Test

- **Does it preserve or enhance human dignity?**

  In purpose, the Aarogya Setu app has been developed to preserve or enhance human dignity. It does so by educating people about the spread and how we can limit the risk of COVID. We can also locate people who are suffering from the virus and identify the near hotspot zones so that everyone in that zone gets tested or can limit their movements for certain weeks. However, it lacks certain controls to make sure that the personal health data has not been breached and is confidential, which may violate human dignity and can cause skepticism.

- **Does it preserve the autonomy of the human?**

  Life-logging technologies capture all the user data on a daily basis. The application here takes the consent of the data being processed but only to a limited threshold. The ToS of this application does not take consent on the anonymized data which can further be used by third parties for research and training.

- **Is the processing necessary and proportionate?**

  The data this application asks for in the initial stages and in the backend is not minimal and necessary for processing. Controls need to be set up to justify the proportionality and requisites of the data.

- **Does it uphold the common good?**

  It is a well-intended application to oppose the spread of COVID in India. With neutral technological capability, it could be argued to the way that it absolutely relies upon the utilization to which it was put.

Recommendations

- **For legality:**

Accountability for Data Governance: There is an absence of certain authority to be held responsible if anything goes wrong with the collected data as no particular controller of data has been identified which lacks accountability. The ToS clearly exempts the Government from any answerability in case of unauthorized access to the user's personal data.

The overall framework of this application could be totally handed over to the public health institutions making them the controller and processor of the data.

The IFF(Internet Freedom Foundation) pointed out the legality to mandate the use of the Aarogya Setu application. According to a review by MIT Technology, "*India is the only democratic nation throughout the world which has made its COVID tracing application mandatory for all the people from the public to private sectors*" **(Jha, 2020)**. It was made compulsory to have this application on our smartphones under the Disaster Management Act. Several questions were raised in the motion. This could have worked if before launching the application, it would have been accredited by the MPs (Members of Parliament) as this issue was of national importance.

There are more and more services being added to the application like e-passes, PM CARE fund transfer and so on which makes it completely irrelevant as the main purpose of this application becomes obscure.

The government (if required) can access the specific masked/anonymized information from the health department for their later administrative or medical interventions rather than taking all the unnecessary information from the application.

- **For proportionality:**

  A detailed specification could be released containing information on the API, cryptography, steps of data anonymization, Bluetooth, and GPS. Currently, the source code of the present version of the application and the source code released are contradicting each other. It will be much better if the server-side code of the application is released.

  There could be a team that can work on the governance and privacy assessment of this application reflecting the current or upcoming risks associated with the process of rollouts.

  There should be a proper adherence towards the clauses in the principal legislation stating that once the COVID 19 pandemic is over, all the data from scratch shall be permanently destroyed from the application as well as the cloud servers assuring the public that these interim interventions will not further be used for any type of scrutiny or surveillance systems.

  Contradictory statements were made by Abhishek Singh, the chief executive of MyGovIndia under which the application was developed. He told " Data will not be used for any other purpose apart from medical emergencies and no third party will have access to the data" **(Sampal, 2020).** Concurrently, as per the new protocols released by the Ministry of Electronics and Information Technology (MeitY), it was clearly mentioned that data will be anonymized when it will be shared with any third party. The government here needs some transparency which should ensure the citizens of what and how their data is shared with proper information privacy documented under some security guidelines.

Summary and Reflection

Data nowadays is directly proportional to the critical decision-making and the strategies in the market. Ask any prominent leader that managing data is still

sometimes challenging for all of them due to no formal guidelines for data governance, no comprehensive data management strategy, complex data sources, and so on.

- This activity has elevated my personal learning to the next level with a primary context of *ethical, legal, and data governance requirements* in different domains particularly the health sector (as in the assignment).
- It helped me understand the *roles and responsibilities of different stakeholders* concerning legislation.
- It also gave me a chance to *understand and critique an application* of my own country, India with the proper guidance of this course structure under **Dr. Emma Murphy**.
- It also helped me interpret what *impact data privacy* is for different organizations and individuals with parallelly *maintaining the quality of the data*.
- It also introduced the importance of *consents*, various *GDPR functionalities*, *pseudonymization*, *First Principles Ethical Test*, *Data lifecycle,* and *Data stewardships.*
- **Key Findings** related to the application can be briefly explained through these points:
    - The Government of India is continuously trying to address the concerns related to the application.
    - This application has the ability to detect the rooted phones and jailbreak but it can be easily bypassed with hack scripts.
    - The older versions had bugs through which an application can easily read inside files, however, it was fixed in the latest version.
    - It tracks the location of the user which on an international scale was unnecessary.
    - The rate of False Positive in contact tracing was more **(Raman et al., 2020).**
    - There were allegations made on the false claims of the government which states no personal information was exchanged once the user registration is done.
    - The efficacy of data is still questionable.

References

Ahmed, N., Michelin, R. A., Xue, W., Ruj, S., Malaney, R., Kanhere, S. S., Seneviratne, A., Hu, W., Janicke, H., & Jha, S. K. (2020). A Survey of COVID-19 Contact Tracing Apps. *IEEE Access*, *8*, 134577–134601. https://doi.org/10.1109/access.2020.3010226

Alderson, E. (2020, May 7). *Aarogya Setu: The story of a failure - Elliot Alderson*. Medium. https://medium.com/@fs0c131y/aarogya-setu-the-story-of-a-failure-3a190a18e34

Dewan, A. (2020, May 10). *Aarogya Setu: A legal and ethical dilemma?* The Financial Express. https://www.financialexpress.com/lifestyle/health/aarogya-setu-a-legal-and-ethical-dilemma/1953985/

Gupta, R., Bedi, M., Goyal, P., Wadhera, S., & Verma, V. (2020). *Analysis of COVID-19 Tracking Tool in India: Case Study of Aarogya Setu Mobile Application. Digital Government: Research and Practice*, *1*(4), 28:1-28:8. https://doi.org/10.1145/3416088

Jha, F. (2020, May 13). *Aarogya Setu scores 2 out of 5 points in MIT review for Covid tracing apps*. ThePrint. https://theprint.in/world/aarogya-setu-scores-2-out-of-5-points-in-mit-review-for-covid-tracing-apps/420746/

Kurian, O. C. (2021a, January 12). *A Leaked COVID-19 Test Result* [Redacted documents randomly selected from the breach]. ORF. https://www.orfonline.org/expert-speak/data-privacy-pandemic-india-just-had-the-biggest-medical-records-breach-ever/

Kurian, O. C. (2021b, January 12). *Data, Privacy, Pandemic: India just had the Biggest Medical Records Breach Ever*. ORF. https://www.orfonline.org/expert-speak/data-privacy-pandemic-india-just-had-the-biggest-medical-records-breach-ever/

Newsclick Report. (2020, September 11). *Tech, Health Activists Want Law to Protect Leakage, Use of Aarogya Setu Data*. NewsClick.

https://www.newsclick.in/Tech-Health-Activists-Law-Protect-Leakage-Use-Aarogya-Setu-Data

Raman, B., Banerjee, S., & Sharma, S. (2020, July 8). *Opinion | On the proportionality of Aarogya Setu*. Mint. https://www.livemint.com/opinion/online-views/covid-19-tracking-app-on-the-proportionality-of-aarogya-setu-11594183812518.html

Sampal, R. (2020, May 1). *User data in Aarogya Setu fully secure, app won't share personal details: MyGovIndia CEO*. ThePrint. https://theprint.in/india/user-data-in-aarogya-setu-fully-secure-app-wont-share-personal-details-mygovindia-ceo/412067/