Demographic Analysis -
A tool to combat Pandemics

Report on Ethical Implications

Lakehead UNIVERSITY

955 Oliver Road Thunder Bay P7B 5E1

Authors

Group 11

1. Chandrasekaran, Dhivya

Student Id: 1111017

2. Ewen, Lisa

Student Id: 0655603

3. Mishra, Devanshu

Student Id: 1104093

4. Nagrecha, Abhishek

Student Id: 1097663

5. Rapelli, Sai Rohit

Student Id: 1107345

6. Wang, Xiang

Student Id: 1106943

7. Zafadiya, Neel

Student Id: 1115533

TABLE OF CONTENTS

Executive Summary:

We propose a data collection project to perform demographic level data analyses which provides a data driven solution to tackle the pandemic. Two major issues with data collection using IoT or technology-based methods is the ability to track the data back to the individual, thus comprising their privacy, as well as the collection of unwanted and irrelevant data. We focus on three major ethical issues: privacy, transparency and informed consent, and security. The data collection process is carried out by volunteers who are provided with clear details of how and why the data collected is being used, thus acquiring informed consent from general public. The data points are made abstract, avoiding personal identification or misinformation ensuring privacy and security. The data collected is then used for data analytics to predict future spread of the pandemic, or suggest data-driven demographic level measures to contain the spread and help plan optimal demographic level health care support.

1. Introduction

Modern technology is equipped with an unfathomably large amount of data which has enabled fields such as data analytics and big data to exist and thrive in the current technological context [1]. Having access to such robust datasets has become especially useful in the climate of global pandemics among the likes of SARS, H1N1, Ebola, and COVID-19 [2]. Having large amounts of data regarding these viruses, as well as future viruses, can enable technology to provide insights such as the locations of active cases, potential variations in symptoms among different areas, analytics on health data to aid with early detection of infection, etc. Despite these incredible opportunities given by the availability of data, the requirement of data collection results in the emergence of new and unique ethical concerns involving privacy, data collection methodology, biases or incompleteness among the data that may skew analytics, among others. In order to determine the fundamental improvements that need to be made with regard to those ethical concerns, it is important to discuss what ethical issues are present in current data collection initiatives.[3]

2. Current situation

When discussing data collection in general, not just in the context of a global pandemic, the ethical concerns are often on behalf of the consumer who is having their data collected. The most common issues that are important to consumers are security, transparency, informed consent, privacy.

One of the most commonly discussed concerns is security. Consumers may consent to have their data collected and utilized for the purposes the application deem appropriate but are often concerned about their data being protected from unauthorized access. As a result of this concern is widely discussed, it is one of the most adequately addressed ethical concerns.

Another concern that has a relatively large body of discussion surrounding it is the concern of privacy. While similar to security, privacy mainly involves maintaining some aspects of a user's data private by not collecting certain data points. Unfortunately, despite the fact that it has become an important point of discussion regarding data collection, there is often little consideration on behalf of most applications to protect a user's privacy. Most commonly a user's data is collected in its entirety, even if not all data is necessary for the processes the application will use the data for.

This leads to the topics of transparency and informed consent; two of the most grossly overlooked ethical considerations on the part of companies who are collecting data. Consent is required for most applications to legally be able to collect a user's data, however, the methods of obtaining a user's consent are often in the form of obtuse and complicated conditions that bypass the user's ability to give informed consent. Most applications are also not forthcoming with what data they are collecting, how they are using it, and if it is being sold to other companies, which violates the important idea of transparency. As informed consent and transparency are aspects of privacy, there has been some discussion regarding

these unethical methods of collecting data, however, very little has resulted in an overall change to data collection methodology.

These ethical concerns in existing data collection endeavors are especially important to address when discussing data collection with respect to the collection of medical data, as the type of data being collected is much more sensitive than internet browsing history or social media interactions. As a result of this sensitivity, ethical concerns will have to comply with bioethics, medical ethics, and HIPAA (USA) and PIPEDA (Canada).

3. Analysis

i. Key Facts

Awareness of the risk of contracting a contagious virus or disease is an essential aspect of aiding the public in maintaining safety measures, and an important goal to achieve this awareness is the development of systematic and ethical data collection mechanisms. It is also important to concentrate on the collective data as opposed to individual data to obtain more regional information that can be obtained from collecting individual data points such as age, gender, existing conditions, members of the household, travel history, etc. This regional data will make it possible to provide the public with information about the potential risk of contracting an illness associated with any given area, and thus, reducing the potential spread of the illness. Once collected, it is also possible to train algorithms to predict a potential outbreak in a given region based on historical and current trends of a given illness.

It is important to note that the data will not consist of any identifying data points such as name, address, etc., and the individual will be responsible for entering their own data through an entry point as part of a web-based dashboard. By not collecting identifying data points and allowing users to input the data themselves will protect user privacy. This data will be stored in an encrypted server location.

ii. Key Stakeholders

The main stakeholders are the users of the dashboard who are both contributing to the data, and viewing the visualizations and analytics the data offers for public health information. It is essential to provide users with an efficient and effective way of inputting their data to avoid incomplete, inaccurate, or poor data, as the success of the dashboard is dependent on the quality and quantity of data available.

iii. Economic Drivers

An obvious economic driver is the COVID-19 pandemic – the fear surrounding a pandemic result in more financial interest in potential solutions to the current and future pandemics. Due to the existing climate

of public health, people will more likely be drawn to a web-based dashboard capable of data collection and visualization that is accessible to a large percentage of people than any other given time. As the design consists of a simple and user-friendly web interface and a secure database, the cost of design is relatively low which makes it an attractive solution during a financially tumultuous time for most.

Another essential, though perhaps more obscure, economic driver to consider is the landscape of visualization tools available for integration with the dashboard. Among the possible options are Google Charts, Tableau, Grafana, Chartist.js, Fusion Charts, Data wrapper, Info gram, Chart Blocks, and D3.js. While some of these options – and others not mentioned here – are open source software, many data visualization tools are proprietary software which require an overhead cost of a licence to use. The ever-changing nature of the open source vs. proprietary software discussion can have a significant impact on any project aiming to make use of data visualization tools.

4. Ethical considerations for our product or service:

As was introduced in section 2, users are often concerned with security, privacy, transparency, and informed consent. To address transparency and informed consent, we are allowing users to enter in their data of their own volition while providing clear insights on how their data will be used via the features available on the dashboard. While collecting data in a manual way is time-consuming and often labour-intensive, it allows us to collect data that is collected ethically and more carefully than data collected through tracking user behaviour. To address the concerns of privacy, personal identifiers will not be collected or considered at any stage of data collection. Having de-identified data loosely addresses the concern of security due to the fact that the collected data presents very little risk to the users should a breach occur in spite of the encryption of the data. By allowing users to enter their data manually, and de-identifying all data collected, the dashboard addresses several ethical concerns related to data collection while also allowing for permanent and data-driven solutions for future needs.

4.a Costs Benefit analysis

i. Costs for Data Collection:

Questions to consider	Estimated costs	Points to remember
- Are our data files, spreadsheets, health records all stored in a uniform format clearly named with unique file names and well organised? - Do we need to ask participants for their consent for data sharing? - Do we collect data on the same level of detail that we will be able to process?	-Data collection, organising the style, formats and asking for consent can be done by volunteers. No cost involved.	- If we plan data organisation beforehand by developing templates and data entry forms for individual data files and by constructing clear file structures, low or no additional cost will apply. - When consent for data sharing is considered as part of standard consent procedures very low or no additional cost apply.

ii. Costs for Data Storage:

Questions to consider	Estimated costs	Points to remember
- How much data storage space is needed for the entire duration of our project? - Do we need to set up a data model and accompanying database for the data? - Do we need a data warehouse?	-The data storage is a cost for the project. We have to consider the factor that this cost is acceptable. - Cloud database as a service: 7 USD/month (storage 5GB, transfer 30GB) [4].	- Institutional storage is often the best way to work on small projects. - Costs for additional storage could include server or disk space, as well as the costs of setup and maintenance. -Reduced IT costs. Moving to cloud computing may reduce the cost of managing and maintaining your IT systems.

iii. Costs for data privacy and security:

Questions to consider	Estimated costs	Points to remember
- Should we protect the data against unauthorised access or disclosure? - Is an institutional server available where we can store your data safely? - Can security be arranged by institutional IT services or is extra software/hardware needed? - Do our data files need encryption before storage or transfer?	-The data would be password protected and only shared between our project team members. There is no need for any additional encryption to be done. Hence no cost involved.	-For confidential or privacy-sensitive data, determining the conditions for controlling access to shared data may require extra time and discussion.

5. Conclusion

With the amount of data currently available, it makes sense to use that data to help fight against public health emergencies such as COVID-19. With that in mind, it is important to acknowledge that not all data would be useful in this endeavour due to the type and cleanliness of the data in question. In addition to that, a lot of available data violates user privacy, is collected without their knowledge, and as such could be considered “unethical data”. In order to overcome the concerns of privacy, ethical data collection, and quality we have proposed a framework incurring minimal costs that allows users to input their own data as part of a global data collection initiative. This framework will consist of a dashboard that allows users to input data without any identifying data points to maintain privacy, and later, have access to the future analytics and other applications of the data such as maps of current or potential future hotspots. The

framework also includes strong encryption on all data to acknowledge security concerns.

At its core, this dashboard provides a pool of data collected by ethical means. The data collected will also be less biased than data collected for monetary gain without the user's knowledge, as the users themselves are choosing to share their data knowingly. Having this data provides the strong foundation necessary for the type of analytics and visualizations that will be useful toward fighting COVID-19 and any other future pandemics.

References

[1] Data Collection, Ethical Issues in. (2020, June 25). Retrieved from https://www.encyclopedia.com/social-sciences/encyclopedias-almanacs-transcripts-and-maps/data-collection-ethical-issues

[2] Disease outbreaks by year. (2020, June 18). Retrieved June 25, 2020, from https://www.who.int/csr/don/archive/year/en/

[3] Hand, D. J. (2018). Aspects of Data Ethics in a Changing World: Where Are We Now? *Big Data, 6*(3), 176-190. doi:10.1089/big.2018.0083

[4]In Search of Free or Cheap Database Storage. (2016, Jan 24). Retrieved from https://medium.com/@bkawk/in-search-of-free-or-cheap-database-storage-d6144ccfa42b

- **Appendix**

Humane Design Guide

Humane technology refers to a technology or design that increases human resolution to problems that humans face. In our project we are using a design guide provided by Center for Humane technology in order to evaluate and identify opportunities for Humane technology. This guide provides a framework to evaluate the involvement of human sensitivity, levels of opportunity to improve related human sensitivities, and steps required to implement improvements. When a human is exposed to a new technology, it often affects one's instincts; these instincts are referred as human sensitivities. The following table contains a list of human sensitivities and the level of opportunity to improve them using our project.

Emotional	Medium
Attention	High
Sensemaking	Medium
Decision-making	High
Social Reasoning	High
Group Dynamics	High

1. In what ways does your product/feature currently engage Human Sensitivities?

The primary focuses of our project are collection, analysis and presentation of data, and can be vulnerable to unnecessary data being exposed to public. The data featured here revolves around a highly contagious illness which, even if presented correctly, may be met with fear or stress. With that in mind, the availability of global data providing insights on the current situation will help improve others' awareness of their potential for exposure and may result in a reduction of infection as a result. This will also aid an individual in

making decisions about travel, both small and large. The aspects of social reasoning and group dynamics are interconnected, and they are completely dependent on individual's social background, however, our project still holds capabilities to impact social reasoning and group dynamics by providing important public health information to society.

2. How might your product/feature support or elevate human sensitivities?

Our project involves data collection and presentation, which makes keeping the presentation of data simple and read easily important. By providing data that can be easily understood among the population there is potential for a reduction in fear or stress, and therefore, supporting emotions. The correct presentation of data will also allow for an increase in awareness which aids in decision-making and social reasoning. In this way, people can better identify their status in society by having correct information about recent trends which helps to improve group dynamics as well.

3. Action Statement

We are providing a dashboard which includes a space to collect information regarding COVID-19 or any other future pandemic, with capabilities to visualize and perform analytics on the collected information. We are providing data about any given disease or illness that may result in worry within infected areas, but it also initiates a reverse effect where infection rates are relatively low. The dashboard will also contain information from authentic organizations about steps to prevent disease, helping to reduces fear and stress within all areas – highly affected or otherwise – thereby improving emotional impact. The data on the dashboard improves attention to disease or other unknown threats which aims to improve decision making. The data about disease from an individual region can help an individual to identify their current situation and can help them improve their social relationship with others among the community, helping to improve social reasoning and group dynamics. While the elevation of different sensitivities will always be unique to a given individual, we believe that in general our proposed design will greatly improve human sensitivities.

1.How is traditional ethical thinking relevant to your product or service?

The traditional Utilitarian approach plays a significant role in our project, as we want our system to produce the greatest good and does the least harm for all affected customers, employees, shareholders, the community, and the environment due to the current COVID-19 pandemic. Ware also considering the Virtue Ethics approach in in order to provide a structured way to promote ideal virtues that aid in societal functions. These virtues are dispositions and habits that will enable us to act according to the highest potential of our character and on behalf of values such as truth, honesty, compassion, generosity, tolerance, fidelity, integrity, fairness, self-control, and prudence.

2. What professional ethics and codes of conducts are important to your product or service?

- For our system the below mentioned ethics and codes of conducts are of utmost importance:

- **Contribute to society and to human well-being, acknowledging that all people are stakeholders in computing.**
  Due to the fact we are collecting required information of as many people as possible, we will be capable of performing data analysis to forecast areas that may require significant attention. Furthermore, our system can contribute in the development of innovative software and applications that are an integral part of the rapid reform of the federal bureaucracy itself.

- **Be honest and trustworthy and Respect privacy.**

  For our product, accountable data use is a desirable implementation, however, it is a challenging endeavour. In order to execute accountable data use, there needs to be institutional reform rather than simply enabling our representatives to defend themselves. There is, therefore, no collection of confidential data about individuals for our project. Only information concerning the diagnosis and study of the spread of the pandemic are issued. Data can include age, gender, present conditions, immediate relatives, etc.

- **Give comprehensive and thorough evaluations of computer systems and their impacts, including analysis of possible risks**

  For our system we intend to explain its complete use and working to each person who might want to understand it. Furthermore, we also intend to publish a manual that could very well explain the idea behind the design of such system and analyse the possible risks.

- **Foster public awareness and understanding of computing, related technologies, and their consequences.**

  Holding data justice in perspective, we strive to integrate the three foundations of data justice. We have proposed a system that collects information from individuals in a decentralised way, with no possibility of discrimination among the data, and allows users to be able to interpret for themselves how their data is used, to whom

they are resold, or the kinds of profiles and interventions those data can enable. These facets ensure a strong engagement with technology.

- **Ensure that the public good is the central concern during all professional computing work.**

  In our case, because we are concentrating on gathering and examining the symptoms of individuals to forecast whether they are exposed to COVID-19, it is very crucial to not overlook the studies from different regions providing information about the way the disease spreads. Ultimately, by considering this work, we can help distribute resources in the best possible way, thus achieving high quality in both the processes and products of professional work.

- **To seek, accept, and offer honest criticism of technical work, to acknowledge and correct errors, and to credit properly the contributions of others;**

  We have proposed to develop our system transparently with open source code, and published privacy model thus creating a room for constructive arguments and debates over the functioning of the system and accepting all our shortcomings

- **To treat fairly all persons and to not engage in acts of discrimination based on race, religion, gender, disability, age, national origin, sexual orientation, gender identity, or gender expression;**

  For eliminating all kinds of prejudice, and to be in sync with the ACM code of ethics[1], we aspire to obtain the data from both the developed as well as developing strata of society in attempt to avoid any discrimination, as well as using methods and techniques easily comprehensible by a layperson.

3. What legal issues, if any, are relevant to your product or service?

Since our system is designed specifically for the United States of America, we have observed the ways in which California laws specifically protect medical information. Under California law, anonymized or de-identified data may not constitute “personal information” as that term is defined by CCPA. Due to this law, we propose to store the data in an anonymized way.

1.https://www.acm.org/code-of-ethics

Moreover, we have also kept the two recently proposed federal bills that focus on COVID-19 data privacy and create new rights for individuals related to COVID-19 health covered entities to: (1) obtain “affirmative express consent” before collecting and using COVID-19 related health information (subject to a few expectations); (2) disclose their data practices related to COVID-19 health information; and (3) create and implement reasonable data security and privacy safeguards. The major privacy concerns that would normally be associated with this type of data collection system appear, on paper, to have been mitigated through affirmative express consent (in the case of the federal bills) [1].

[1] Nick A, Ffion F, Stella C, Chris C, Marcus E, Anna G, Julie H, Tricia H, David K, Christoph R. “Contact tracing apps: A new world for data privacy”, June 2020, https://www.nortonrosefulbright.com/en-us/knowledge/publications/d7a9a296/contact-tracing-apps-a-new-world-for-data-privacy