Data Governance Network
Anchored by IDFC Institute

June 2020

Working Paper 08

Using GIS for public health and its privacy implications: A case study of Kerala

Harsh Vardhan Pachisia, Girish Gupta and Kadambari Shah

IDFC Institute

Data Governance Network

The Data Governance Network (DGN) is a multi-stakeholder community of researchers tackling India's next policy frontiers: data-enabled policymaking and the digital economy. Some of India's leading policy think-tanks – IDFC Institute, National Institute of Public Finance and Policy (NIPFP), Internet Democracy project (IDP) and IT for Change (ITfC) are research nodes of DGN. IDFC Institute also functions as the secretariat of DGN. DGN works to cultivate and communicate research stemming from diverse viewpoints on market regulation, information privacy and digital rights, in an attempt to generate balanced and networked perspectives on data governance—thereby helping governments make smart policy choices which advance the empowerment and protection of individuals in today's data-rich environment.

About Us

IDFC Institute has been set up as a research-focused think/do tank to investigate the political, economic and spatial dimensions of India's ongoing transition from a low-income, state-led country to a prosperous market-based economy. We provide in-depth, actionable research and recommendations that are grounded in a contextual understanding of the political economy of execution. Our work rests on three pillars – 'State and the Citizen', 'Strengthening Institutions', and 'Urbanisation'. The State and the Citizen pillar covers the design and delivery of public goods, ranging from healthcare and infrastructure to a robust data protection regime. The Strengthening Institutions pillar focuses on improving the functioning and responsiveness of institutions. Finally, the Urbanisation pillar focuses on the historic transformation of India from a primarily rural to largely urban country. All our research, papers, databases, and recommendations are in the public domain and freely accessible through www.idfcinstitute.org.

Terms of Use

Disclaimer

The views and opinions expressed in this paper are those of the authors and do not necessarily represent those of the organization.

Suggested Citation

Pachisia, H. V., Gupta, G. and Shah, K. (2020). Using GIS for public health and its privacy implications: A case study of Kerala. *Data Governance Network Working Paper 08*.

Supported by a grant from Omidyar Network

Abstract

Governments, globally, are turning to data and technology to improve governance, specifically service delivery, with tools such as digital healthcare sensors and the Internet of Things. At the same time, these technologies are fusing into the lives of their users; the ballooning of personal data has sparked a conversation on privacy and safety. With a focus on Geographic Information Systems (GIS), which can greatly enhance planning and public service delivery, we look at privacy issues surrounding health data. We highlight GIS and health, not just in theory, but by bringing in lessons learned from our ongoing work using GIS to improve service delivery in Kerala. We document our takeaways to help future deployments of the technology, both in Kerala and in other Indian states. Recommendations are provided on how Kerala should use GIS to improve its healthcare delivery without infringing on the privacy of its citizens.

Table of Contents

1. Introduction

Technologies such as Artificial Intelligence, digital healthcare sensors and the Internet of Things are fusing into the physical lives of their users. Data and technology are being employed by the private sector as well as governments, playing a central role in the planning and monitoring activities of state performance and in the design of e-governance platforms. In India, several states, such as Andhra Pradesh, have developed dashboards across various infrastructure sectors which are crucial for real-time monitoring. At the same time, as governments are turning to technologies to improve governance, the ballooning of personal data has rightly sparked a conversation on safety. However, there are prominent technologies that can be advantageous to governments and communities without infringing upon personal privacy. One is the Geographic Information System (GIS). GIS makes extensive use of non-personal data and can have several benefits, particularly in improving local planning and public service delivery.

IDFC Institute has been invited by the Government of Kerala to assist in the deployment of GIS across the state to help improve service delivery, while ensuring that the personal privacy of their citizens is not affected. Our work with the state is two-fold:

1. Technical: The use of technology is critical in building out an effective GIS system, drawing on the latest means of tech-based data collection and widespread use of task management applications. To that end, we are building out a campaign management and analytics software to manage the data collected as well as developing user-interface guidelines.
2. On-the-ground realities: We are also creating and delivering training modules, instructional manuals and videos for different stakeholders in government and for the citizens of Kerala on using GIS-based systems.

Documenting the challenges faced and the lessons learned through these efforts will help us understand how an effective GIS-based system can be built, deployed and scaled. With that in mind, this paper aims to analyse two research questions:

1. Why should policymakers in Kerala use GIS to improve public service delivery (with an emphasis on healthcare) without infringing on the privacy of the state's citizens?
2. What are the on-ground steps and challenges that need to be overcome to build out an effective GIS system for the state of Kerala?

To demonstrate why the government should employ GIS, we first outline the theoretical aspects and use cases of non-personal data, specifically focusing on GIS. Next, we describe how GIS can be used to improve service delivery, with an emphasis on the healthcare sector. Third, we highlight current privacy regulations, their relation to GIS and the pros and cons of different data anonymisation techniques to ensure that the usage of GIS does not infringe upon personal privacy. We then summarise the steps we have taken on-ground, along with the challenges and lessons we have learned in Kerala. We analyse how the government can make use of GIS to augment service delivery of healthcare. Finally, we provide a list of policy recommendations for both Kerala and other state governments interested in using GIS to improve governance, both in the healthcare sector and otherwise.

2. What is Non-Personal Data and why is it important?

Rapid innovation has resulted in pervasive technological and social change[1]. It has given India the Aadhaar system and the United Payments Interface, and the world WhatsApp and Facebook. These platforms have fundamentally changed the functioning of business, governance and society; every individual and firm leaves behind a large trail of data. As a result, India has seen a wide-ranging debate on the safety of personal

1 Schwab, K. (2017)

information that is part of such data trails; this is understandable given that extensive use of sensitive and Personally Identifiable Information (PII) has led to concerns about user privacy. Consequently, the regulation of these products dominates policy conversations. Crucial measures such as the Personal Data Protection bill (PDP), the National Digital Health Blueprint (NDHB) and the proposed formation of the Data Protection Authority are on the anvil. In particular, the PDP bill defines personal data as "data about or relating to a natural person who is directly or indirectly identifiable, having regard to any characteristic, trait, attribute or any other feature of the identity of such a natural person".

While a lot of emphasis has been placed on the regulation of personal data in India, little, if any, attention has been given to new technologies and data points that do not require PII to benefit society and will not be affected by potential new regulations on the use of personal data. These, we think, should be promoted. Such information falls under the ambit of Non-Personal Data (NPD), defined by the PDP bill as "any data other than personal data". This definition provides little clarity on the landscape and understates the massive importance of non-personal data. On the other hand, the European Union (EU)'s General Data Protection Regulation (GDPR) defines NPD as data which originally did not relate to an identifiable natural person or information which was initially personal data, but later made anonymous[2].

In addition to not identifying a particular individual, NPDs usually carry immense economic value since they provide insights based on aggregate patterns. For example, insights on commuter patterns can be revealed from traffic data. Transport for London, the governing body for public transit in London, is estimated to generate over USD 160 million annually by providing access to its geospatial datasets. Geospatial data, in general, was shown to have generated over USD 400 billion annually in global revenue in 2017[3]. Other possible uses of NPD include discerning purchasing patterns in a demographic from e-commerce data or making inferences about the spread of diseases based on public health information.

2 Forge, S. (2018)
3 Open Data Institute (2018)

2.1. Types of Non-Personal Data

As can be seen by the PDP bill, the (re-)identifiability of a natural person appears to be at the core of the definition of personal data, both in India and the EU. However, every non-personal data point does not have the same impact on a person's privacy. With that in mind, NPD should also be split into different types based on the privacy risk posed. Singh, et al. (2019) outline two main categories of NPD:

1. Non-human NPD: This includes data that did not originate from and cannot identify a human being. Examples of such data include macro-level data such as weather forecasts and Gross Domestic Product (GDP). It could also include data on supply chains, aggregated e-commerce sales, road networks, etc. Further, non-human NPD comprises aggregated data from multiple individuals where individuals are not identifiable. For instance, the frequency and number of passengers carried by public transport and commute patterns (such as those provided by Uber Movement[4]) fall under this category.
2. Human NPD: This category includes data that originally identified or pertained to individuals but has subsequently been anonymised, ideally making it extremely difficult to identify the underlying people. This category includes anonymised datasets that contain personal data such as e-commerce shopping histories, location data, personal health records, etc.

The usage and flow of NPD is currently not regulated in India and does not fall under the purview of the PDP bill. However, the economic value generated by NPD has led to calls for its regulation. For instance, the 2018-19 Economic Survey of India recognised the value of such data, classifying it as "a public good which may be utilised for the economic benefit of the country"[5]. Such calls have resulted in the country's nodal ministry for information technology — the Ministry of Electronics and Information Technology — constituting a Committee to deliberate on these very concerns and formulate a data governance framework for non-personal data.

4 Uber releases limited, publicly accessible data through Uber Movement, which provides average travel times between two zones for a given time period across a city.

5 Krishnan, V. (2019)

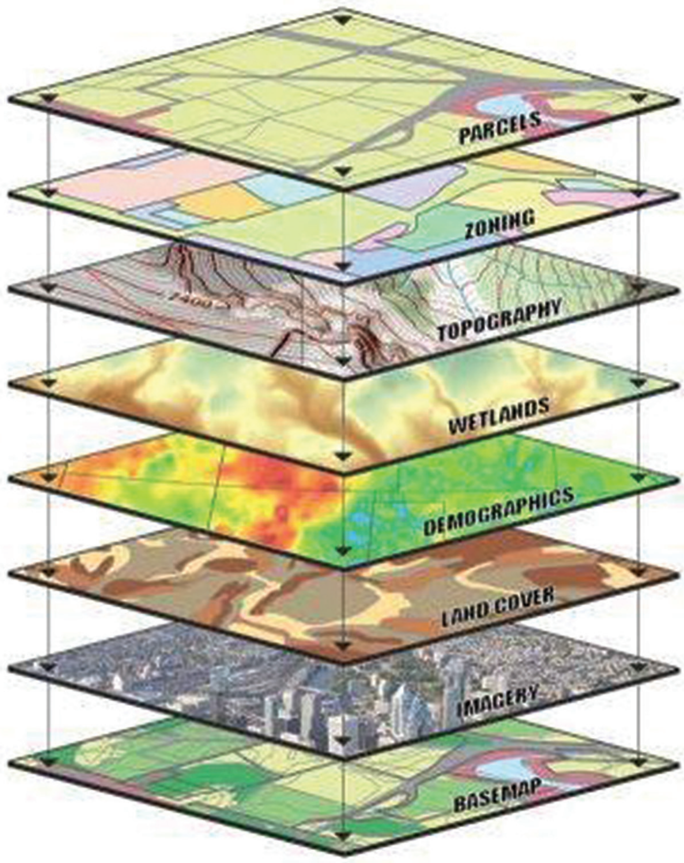
3. Understanding GIS: A non-personal data using system

3.1. What is GIS?

GIS is a system that can capture, store, and analyse geo-spatial or spatially-referenced data. In particular, it allows spatial data to be connected with non-spatial data points. This can be achieved through a variety of keys — for instance, latitudes and longitudes to connect disparate datasets such as weather and administrative boundaries. Combinations like these allow for spatial analysis of otherwise non-spatial data, making the use of GIS extraordinarily powerful. Another distinguishing feature of the system is its ability to conduct a temporal analysis of multiple spatial and non-spatial datasets.

A GIS database can be built to provide insights by 'layering' different information on top of each other. Each 'layer' (Figure 1) in a GIS database can either contain

Figure 1: A representation of layers in GIS



Source: Kolios, S., Vorobev, A.V., Vorobeva, G.R., Stylios, C. (2017)

raw/non-human NPD, such as satellite imagery and weather data, or thematic/ human NPD, such as aggregated data on health services provided.

As outlined by Aghajani, et al. (2017), a GIS structure requires five main components to be effective:

1. Spatially-referenced datasets containing both non-human NPD (including road networks and zoning information) and human NPD (aggregated personal health records, commute patterns of a region and so on).
2. Hardware tools including a data collection mechanism (such as smartphones), a data streaming service using an Application Programming Interface (API) and a server that physically stores the data.
3. Software by which users can access, query and analyse the database (for example, QGIS, ArcGIS, etc.).
4. Data sharing and data management procedures between different data collecting stakeholders.
5. Qualified people to collect on-ground data as well as process and analyse the database to provide recommendations.

3.2. Technical evolution of GIS

The fundamental principle of using geographical information to provide insights and aid decision making in sectors (for example, healthcare and education) was present even before technologies such as GIS were developed. Hippocrates, in the 3rd Century BC, in *On Air, Water and Places*, described the role of weather and places in medical treatment outcomes. The French geographer Charles Picquet developed the first colour-coded maps of cholera distribution in 1832 Paris. 18 years later, John Snow became the first epidemiologist to find clusters of the cholera outbreak that gripped London: by using a dot map and analysing sufferers' locations, he was able to identify the source of the outbreak—a water pump on then Broad Street[6].

6 Aghajani, J., Farnia, P., & Velayati, A. A. (2017)

While such preliminary uses of geographic information to aid decision making were useful, the time-consuming nature of drawing maps by hand and collecting raw data was a major deterrent to mainstream adoption. This changed in the 1960s when researchers in the United States developed the first GIS that allowed for computer mapping and analysis. This initial system was used extensively for resource assessment, planning and land evaluation. Over the next two decades, local US government agencies regularly started adopting GIS to experiment with new policies and aid decision making. By the end of the 1980s, the private sector had commercialised the software with many companies offering GIS-based services. In the 21st century, the amount of data that can be mapped onto GIS has mushroomed, in large part due to innovations such as smartphones, whose hardware and software have become more sophisticated, as well as a growing understanding of the power of data. The field now encompasses a wide variety of GIS-based applications that are used in agriculture, education and healthcare planning, botany and zoology. Further, this system can be used to provide users important analytical insights for good governance, which can be applied to sectors such as urban planning and disaster management[7].

4. Using GIS to improve public service delivery

The Indian state is highly understaffed with the least number of bureaucrats per capita compared to its G20 counterparts[8]. As a result of such scarcity, the planning for and delivery of public services suffer, posing multiple challenges for the country's developmental activities. While the Indian government will have to increase the number of personnel in the longer-term, the usage of GIS can help optimise and effectively utilise the limited personnel currently at the state's disposal when strategising local planning and organising delivery of services. For example, in terms of planning, watershed management requires complex calculations over huge amounts of topographical, hydrological and weather data. GIS-based systems can make the computation easier, automate significant portions of the process and reduce the need for specially trained hydrologists. At the same time, for service delivery, GIS can help policymakers accurately decide

7 Grant, A., Razdan, R., Shang, T. (2017)
8 Ibid.

where best to place schools, banks and primary healthcare centres. Following such an analysis, they can deploy their limited personnel to these spots where their efforts will have the maximum impact.

The use of GIS-based systems can also help reduce knowledge gaps by placing multiple layers of data from different departments onto an integrated platform, such as from the police, electricity department, telecom ministry, etc. Such a platform would assist policymakers in making more informed local level plans. Furthermore, if such a platform is made public with the requisite privacy safeguards, it can form the foundation for creating and giving public access to useful data services. This, in turn, would increase public engagement through transparency and promote innovation.

These benefits of using GIS-based systems cut across all sectors. However, there are also specific ways in which GIS is employed by different sectors to improve service delivery. Below, we briefly outline these specific use cases before focusing on how GIS can help improve healthcare delivery.

4.1. Managing and planning for natural disasters

GIS and open mapping technologies have been used across the world by governments, aid agencies and the private sector. The idea is that the base maps are utilised across departments and agencies, providing a common platform for authorities to prepare for natural disasters, i.e., use comprehensive maps to identify risk-prone areas at high resolution, such as flood forecasting once disasters strike or recognising and providing relief to damaged buildings, roads and general infrastructure.

In mitigating the effects of natural disasters, the American Red Cross, for example, encourages the use of GIS to better analyse hazards and improve the response capacity in communities by advocating their needs to local governments. These insights have proven useful in the past, during and after disasters, for seeking out and targeting the most at-risk, least covered communities[9].

9 Geographic Information Systems (GIS) Use Cases, American Red Cross.

4.2. Law enforcement

GIS has been incredibly useful in law enforcement. The use of GIS and different technologies is critical to ensure public safety and increase the effectiveness of law enforcement with finite resources. For example, Chiemeka, et al. (2019) conducted a crime mapping exercise to help police in the Nigerian city of Owerri. Such a system, combining data from the national crime databases, population data and points of interest such as shopping centres, police stations, hospitals, etc., allowed authorities to identify crime hotspots and detect criminal activity[10].

4.3. Education

The use of GIS has also immensely aided education planning, particularly in choosing locations for new schools. The technology can be used to assess the demand-supply situation as well as map out unreached or underserved areas by combining layers containing information on the locations of existing schools with the number of students served. In addition, to trace travel time for students, one can track on-road distances to schools by adding a layer of road networks.

5. GIS and Healthcare

5.1. Why healthcare?

The rewards of using GIS to improve healthcare planning at the national, state and local level are enormous: policymakers can get better health outcomes, prevent and control the spread of diseases and provide healthcare services to previously underserved communities. However, for the system to be most effective, human NPDs such as aggregated personal health records are critical, the use of which poses a privacy risk. For instance, strategically placing AIDS specialists in districts with a large number of people with the disease would improve health outcomes. However, the security

10 Chiemeka, E. C., Chukwudebe, G., & Ekwonwune, E. N. (2019)

and anonymisation of this information is crucial to avoid any possible community discrimination against AIDS-infected patients. Given such a high risk/high reward scenario, developing privacy guidelines for the use of GIS in healthcare is crucial. By doing so, we can maximise the improvement in service delivery by policymakers while minimising privacy concerns of citizens.

The 17 Sustainable Development Goals (SDGs) are targets set by the United Nations in different sectors for ensuring a better future for the world. Of them, "Ensuring healthy lives and promoting well-being for all at all ages" is critical[11]. Robin, et. al. (2019) state four factors that need to be met to achieve this goal at the local level:

1. Availability: The proportion of doctors, Public Health Centres (PHCs) and skilled service providers relative to the population they serve.
2. Accessibility: The distance between the service delivery points such as hospitals, PHCs, doctors, etc. and the population they serve.
3. Quality: The number of service points that are qualified to properly serve the population.
4. Utilisation: The amount of the population that is aware of and is using such services.

The inability to meet all four factors of this framework at a local level is often the challenge in achieving this SDG. This is where a GIS-based system is particularly powerful; it can address inequalities in geographic access and gauge the availability to PHCs, hospitals, and other medical centres. It can do this by charting out the point data of every healthcare provider as well as its distance to the population of the district. However, for it to be truly effective, quality and actual utilisation of the healthcare providers is critical. This would require the use of certain aggregated personal health records, such as the number of patients admitted and types of diseases treated. To ensure data privacy, appropriate anonymisation techniques need to be applied to aggregate such data, thereby providing information on the factors of quality and utilisation. Ultimately, if information is provided on all the four factors, the use of GIS coupled

11 UN Sustainable Development Goals 2030.

with anonymised and aggregated personal health records can be extremely potent in improving healthcare.

5.2. Use cases

Robin, et al. (2019) applied their four-factor framework and developed a GIS for the rural Habiganj district in Bangladesh to improve healthcare planning and resource allocation. They used spatial imagery and physically mapped out the locations of PHCs, road networks, etc. to understand the underserved parts of the district in terms of access (i.e., distance to PHCs) and availability of services (i.e., number of healthcare providers). Next, they collected data on the number of households present in different parts of the district, types of disease present, locations of the households where neonatal deaths occurred, etc. Following this, they anonymised and aggregated this personal data to remove personal information and identify clusters. Finally, they placed their datasets onto a GIS to identify underserved areas. Further, using the data that was collected on diseases, they were able to prioritise areas where the quality of services provided was inadequate. Ultimately, such an analysis allowed local level policymakers to strategically change budget allocations to different parts of the district depending on its health requirements.

Lawal and Anyiyam (2019) also examined the impact of geographic access to PHCs in Nigeria with a combination of open data and geospatial analysis. In particular, they used only non-human NPD such as data on elevation, location of healthcare facilities, population density and network data. They defined three levels of access to healthcare: good (less than 30 minutes), poor (31 to 60 minutes) and very poor (greater than 60 minutes). They found that healthcare facilities often clustered at major urban agglomeration and transit routes leading to a city. Further, areas with low accessibility were often highly correlated with low quality road networks. Improving such factors could lead to better health outcomes.

In addition to aiding healthcare planning, GIS has been used to predict and track the spread of diseases around the world. This is due to the fact that most causes of diseases spreading are spatial in nature, i.e., their distribution and concentration vary depending on their location. For example, Canada used GIS to monitor the spread of the

West Nile virus[12]. Similarly, in Iran, Mesgari and Masoomi (2009) used GIS to develop a cancer decision support system for policymakers using a combination of geospatial data and data related to death counts caused by cancer in the country. In 2002, Ayuthaya, a city in Thailand, used GIS for mapping disease distribution and examining the effects of different factors on public health. Their data sources included that of population density and the locations at which infectious diseases were prevalent. In particular, the Thai policymakers wanted to understand the link between time and the spread of a disease. Using a combination of spatial and statistical regression techniques, they found that pneumonia had the highest dependence coefficient with time (94%). This analysis surfaced the diseases which required pre-emptive and proactive monitoring by policymakers given their potential to spread and infect an entire population quickly[13].

Each of these examples illustrates how GIS can be employed to improve healthcare delivery in the state of Kerala. In terms of allocating limited personnel, the examples above were from countries that have often understaffed bureaucracies; GIS allowed policymakers in Bangladesh to allocate limited budgets and their personnel to places where they would have the most impact. Further, decision makers in Nigeria were able to understand where best to place new healthcare facilities to ensure easy accessibility and availability to the wider population. These cases also combined datasets from different departments onto an integrated GIS platform, helping policymakers make better, more informed decisions.

6. GIS and Privacy

While GIS has improved—and has the potential to enhance—public service delivery in multiple sectors, there are privacy concerns that need to be understood, especially in the delivery of healthcare. On one hand, GIS programmes using only non-human NPD carry no privacy risks since they do not identify particular individuals (and still hold immense economic value). At the same time, the system is at its most effective when

12 Gosselin P, Lebel G, Rivest S, et al. (2005)
13 Keola, S. (2002)

non-human NPD (such as weather data, road networks and zoning information) are overlaid with human NPD (such as personal health records). Developing and adopting privacy guidelines will be crucial for GIS to be effectively used by policymakers to improve service delivery in Kerala.

Should the public be given access to sufferers' precise locations? If that data were somehow masked to protect identities, would it still be as valuable in finding today's Broad Street water pump? Computing power and the amount of data available have significantly increased in recent decades[14]. This means that masking can more easily be overcome. However, is that outweighed by the health benefits?

These are questions that anyone who collects data must face when thinking about how to make it public.

This trade-off is very different in the health sector as compared to others, where the consequences can be so severe and personal. Few would favour more deaths over an abstract privacy concern. The recent novel Coronavirus (2019-nCoV) pandemic has highlighted this, with many arguing that, given the severity of the potential consequences, insistence on privacy should be relaxed.

Does the benefit of knowing if you have been in close contact with a sufferer of the virus outweigh that person's right to privacy—and does the community's right to know whether you have been near a sufferer outweigh your own right to privacy[15]? As with all security and privacy concerns, there is a compromise to be made between the benefit of open data and the cost of revealing personal information. For instance, knowing who has been infected where can help isolate disease outbreaks and ensure that resources reach those who most need them.

China has gone further than most in executing this, its surveillance system forced somewhat out of the shadows, according to a Reuters report[16]. Cameras coupled with

14 Elwood, S., & Leszczynski, A. (2011)
15 Thorbecke, C. (2020)
16 Yang, Y., Zhu, J. (2020)

machine learning software scanned the streets for people with fevers, recognising their faces and reporting them to authorities who could triangulate the data to find others who may have been affected—and go on to track them. Software could then tell end users if they had been on a flight or a train with a known coronavirus carrier, and maps showed the locations of buildings where infected patients lived. The Chinese government launched a mobile "close contact detector" application[17], which required government-issued identification and phone numbers.

One worry is that the relaxing of privacy concerns is very hard to unravel once a crisis ends. Analogies are being made to the United States and much of Europe after the September 11, 2001 attacks. In many realms, privacy was given up in favour of security[18]. "Has history ever shown that once the government has surveillance tools, it will maintain modesty and caution when using them?" wrote Hu Yong, a professor at Peking University's School of Journalism and Communication[19].

However, these fears are far removed from the work being done in Kerala and so some, at least, can be ignored.

6.1. Current regulations

U.S. law requires healthcare providers to adhere to the Health Insurance Portability and Accountability Act (HIPAA)[20] of 1996, which limits disclosure of patients' data and requires that it be stored and transmitted securely. Critics see this as irrelevant to modern-day medicine where data from web searches, social media posts, grocery shopping, wearable tech and television habits—which are not covered by HIPAA—can reveal far more about a person's health than their medical records. Not only that, the Act can often hamper medical research rather than bolster it. For example, academics trying to answer a simple question about ER treatment would have to jump through

17 Xinhua (2020)
18 Ross, C. (2020)
19 Hao, K. (2020)
20 HHS, US Government (2020)

many hoops to get to data; it has also become a means by which professionals can simply say no to requests at a whim[21].

When bringing in GIS, wider, non-healthcare-specific legislation becomes important. GIS data can provide information such as addresses and unique identifiers that can be used as keys to relate databases and recognise users. Those executing the U.S. census, for example, must not "make any publication whereby the data furnished by any particular establishment or individual...can be identified"[22]. Personally-identifying information is not made public in census results. Names, ages, personal addresses, etc., are all removed in order to "anonymise" the data.

In India, the policy approach to GIS regulation has often been cumbersome and unplanned with a focus on the security concerns surrounding the use of geospatial data. As a result, geospatial data has usually been kept internal by multiple government agencies. For instance, the National Map Policy (2005) made the Survey of India (SOI) the nodal government agency in charge of all processes dealing with geospatial data. As stated by Garg (2016), "while harping for open access to geospatial data, the 2005 National Map Policy accompanied by corresponding guidelines have largely restricted the power to map geospatial data to SOI". In 2011, the Remote Sensing Data Policy made a further distinction between terrestrial mapping and geospatial imagery with the National Remote Sensing Centre (at the Indian Space Research Organisation) becoming the nodal agency for geospatial imagery and SOI retaining charge of the terrestrial mapping.

An attack on an Indian air force base in January 2016 led to the politically-charged draft of the Geospatial Information Regulation Bill, 2016 (GIRB) due to information that spatial imagery had played a key role in the planning of the attack. This bill included proposals such as introducing a license-raj system for the production of geospatial data and placing severe restrictions on the dissemination of geospatial data. Such proposals were met with heavy resistance from India's geospatial industry as well as by public comments from experts recommending that the bill be withdrawn in its entirety

21 Chen, A. (2020)
22 Census Bureau, US Government (2020)

due to multiple harms caused[23]. As a result, the draft GIRB has since been put on the backburner [24].

While the Indian government has taken a scattered and unplanned approach to the regulation of spatial imagery, both national and state governments have instituted multiple initiatives that employ GIS-based tools, in particular to map urban settlements. In 1992, the Urban Mapping Scheme aimed to produce pan-India maps at a 1:25,000 scale using aerial photography for 53 towns[25]. In 2001, the National Capital Region used space-based geospatial technology to prepare its Regional Plan[26]. In 1996, the National Informatics Centre established[27] a remote sensing and GIS division, tasked to establish a multi-layer GIS for planning. In 2011, the first all-India mapping service, Bharatmaps/NICMAPS was launched. The Digital India scheme's vision was to use GIS as a foundation for[28]:

- Digital Infrastructure as a Utility to Every Citizen
- Governance & Service on Demand
- Digital Empowerment of Citizen

By 2010, more than 200 cities in Maharashtra, Gujarat, Karnataka, Rajasthan, Orissa and Andhra Pradesh were using GIS-based planning tools at the level of Urban Local Bodies[29]. The Tirupati Urban Development Authority in Andhra Pradesh commissioned the Environmental Planning Collaborative to prepare a master plan for the region and a zonal development plan for Tirupati town in August 2001[30]. Similarly, the Hyderabad Urban Development Authority, Telangana, used GIS to create a register of land uses as part of its Plan for Sustainable Development, Hyderabad 2020 Draft Master Plan for Hyderabad Metropolitan jurisdiction area[31].

23 Prakash, P. (2016)
24 Srivas, A. (2017)
25 Sub-Scheme on Formulation of GIS based Master Plans for 500 AMRUT cities.
26 Ibid.
27 MapService (1996)
28 Chandra (2016)
29 Sub-Scheme on Formulation of GIS based Master Plans for 500 AMRUT cities.
30 Ballany, S., & Nair, B. (2002)
31 Venugopala, Rao K., Ramesh, B., Bhavani, SVL., & Kamini, J. (2010)

At the central level, the Ministry of Housing and Urban Affairs' Atal Mission for Rejuvenation and Urban Transformation aimed to "develop common digital geo-referenced base maps and land use maps using GIS across 500 cities[32]." The central government also launched the National Urban Information System to generate hierarchical urban geospatial and non-spatial attributes of multiple scales to help in planning and decision support[33].

GIS-based systems have been employed for certain purposes by the Indian government. However, the policy regulations mentioned demonstrate the scattered and unplanned approach to the use of spatial imagery in India, with multiple regulations and government agencies inhibiting the further uptake of GIS-based systems and innovation in the use of such a technology. As we assist the deployment of GIS in Kerala, such regulations need to be kept in mind when building the system. Furthermore, as the Kerala government is using GIS on an open platform, it must install its own thought-out policies to counter potential attacks.

6.2. Means of attack

There are several types of attack. They begin with a simple breach through a hardware or software security hole or simply acquiring data that was not properly masked. Serious security protocols must be in place for the storage and transmission of data, including encryption, hashing as well as strict rules on who has access to the most sensitive data and how that access is obtained.

Things get even more complicated when attackers begin triangulating data.

Imagine a small census tract with a population of just two people. That population figure—two— is made public alongside the average age of the inhabitants of the tract, say 28. There is no further information about the people in the tract available through the census. It would be impossible to ascertain either member of the population's age

32 Sub-Scheme on Formulation of GIS based Master Plans for 500 AMRUT cities.
33 Venugopala, Rao K., Ramesh, B., Bhavani, SVL., & Kamini, J. (2010)

as there are too many unknown quantities and so an infinite number of answers to the question of how old either member of the population is. A local newspaper publishes a story about one member of the tract and identifies the person's age. The other person's age is now very easily calculated.

This is a stripped down and unrealistic example that can easily be applied to any dataset. However, the principle applies to far more complex situations, which essentially could allow people to solve simultaneous equations by brute force to identify people and link them to confidential and sensitive data.

Consider two data sets: medical data and a voter list. The former has had any uniquely identifying information — social security number and name — scrubbed. They are shown below in Figures 2 and 3[34].)

The fields in both data sets, i.e., date of birth, sex and zip code, are quasi-identifiers which allow reconstruction of part or all of the data that publishers had wanted anonymised; this is demonstrated in Figure 4. Variables such as health issue and party are considered sensitive.

Figure 2: Medical Data released as anonymous

Medical Data released as anonymous

	SSN	Name	Ethnicity	Date Of Birth	Sex	ZIP	Marital Status	Problem
1			black	09/27/64	male	02139	divorced	obesity
2			black	09/30/64	male	02139	divorced	hypertension
3			black	04/18/64	male	02139	married	chest pain
4			black	04/15/64	male	02139	married	chest pain
• 5			black	09/15/64	male	02138	married	shortness of breath
6			caucasian	03/13/63	male	02141	married	hypertension
7			caucasian	03/18/63	male	02141	married	shortness of breath
8			caucasian	09/13/64	female	02138	married	shortness of breath
9			caucasian	09/07/64	female	02138	married	obesity
10			caucasian	05/14/61	female	02138	single	chest pain
11			caucasian	05/08/61	female	02138	single	obesity

Source: Sweeney, L. (2002)

Figure 3: Voter list

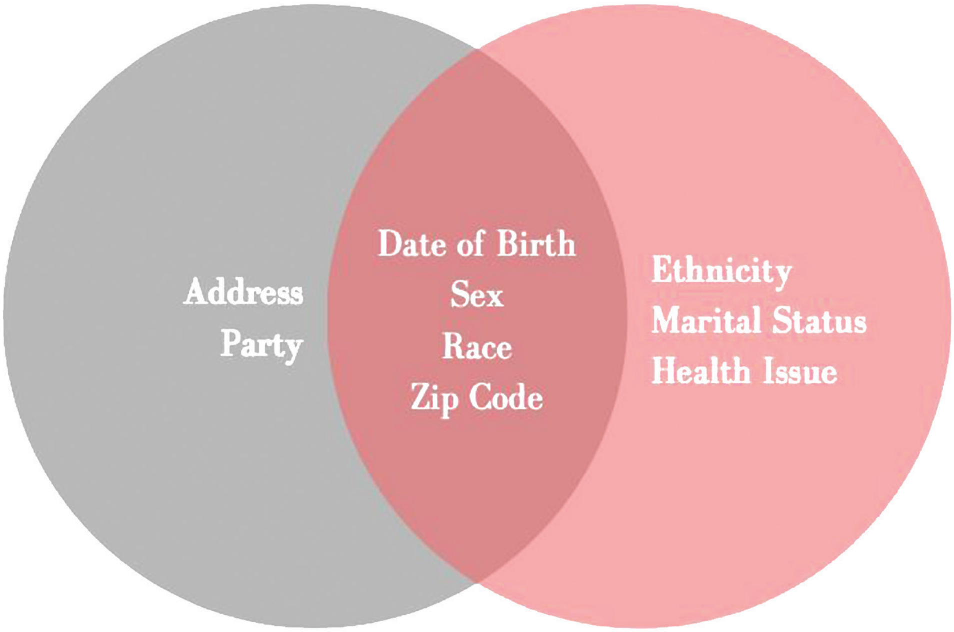
Voter List

	Name	Address	City	ZIP	DOB	Sex	Race	Party	....
	....	....	....	....	....	....	....	....	....
	....	....	....	....	....	....	....	....	....
•	Jim A. Cosby	570, Laurel St.	Cambridge	*02138*	*9/15/64*	*male*	*black*	democrat	....
	....	....	....	....	....	....	....	....	....

Source: Sweeney, L. (2002)

34 Sweeney, L. (2002)

Figure 4: Sample data indicators



Source: IDFC Institute

Jim A. Cosby is male, black and was born on September 15, 1964 according to voter data. Only one row of the medical data, number 5, matches those attributes and allows us to garner further information about Cosby: he is married and has shortness of breath. If not all of those quasi-identifiers were available, Cosby's zip code/address could perhaps have been used to make the link.

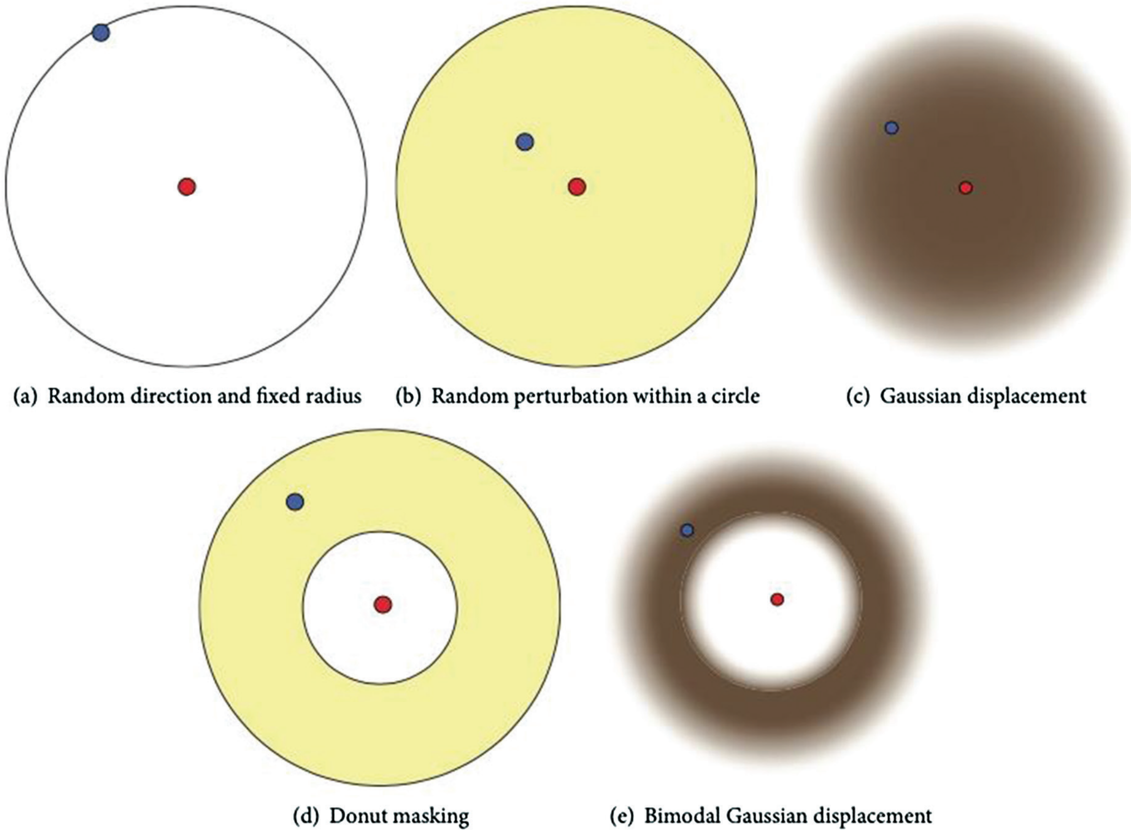
A concrete breach example using GIS and health data took place after Hurricane Katrina in 2005. Researchers were able to re-engineer individual addresses of victims despite data having been effectively blurred to a radius of a block and a half[35]. This showed how easy the process can be even when researchers have taken steps to anonymise data.

6.3. Means of anonymisation

More complex techniques than simply blurring, averaging or suppressing data have been developed over the years. A subset of suppression, the so-called 'rule of three',

35 Curtis, A. J., Mills, J. W., & Leitner, M. (2006)

Figure 5: Example of jittering



Source: Zandbergen, P. A. (2014)

which prohibits publishing of data with fewer than three participants, would prevent some of the issues described above[36]. Another is aggregation, which is a compromise on precision[37].

Another method, more relevant to GIS specifically, includes injecting noise, which, in GIS, is often known as 'jittering'. Rather than publish precise locations, the locations are randomised within a fixed radius, on the circumference of a circle or with more complex variations on that theme as per Figure 5.

It is widely agreed upon that the amount of displacement in a jitter should be inversely proportional to the population density around the point. This makes intuitive sense in that if there are more people in the area, a small possible area will encapsulate enough of them so as to make identifying one person difficult; conversely, if the area is sparsely

36 Garfinkel, S., Abowd, J. M., & Martindale, C. (2019)
37 Armstrong, M. P., Rushton, G., & Zimmerman, D. L. (1999)

populated, a larger possible area will be required in order to encompass enough people so as to make identifying one person difficult. This is especially important if a large part of the area chosen covers, for example, a lake or forest in which presumably no one resides.

Using this method could have prevented researchers pinpointing hurricane victims' home addresses and can possibly be applied by policymakers in Kerala.

6.4. Measures of anonymisation

Measuring the effectiveness of anonymising data involves working out the probability of discovering all or part of an anonymised dataset.

One measure is *k-anonymity*, proposed by Samarati and Sweeney (1998). A dataset is *k-anonymous* if there is a minimum of *k-1* people with the same characteristics when anonymisation has been completed. If *k* is not sufficiently high, more anonymisation is necessary.

Using the techniques to anonymise data outlined above—in this case, generalising the zip code, sex or date of birth—could turn the data in Figures 2 and 3 into a dataset whereby it would be impossible to link more than *k* people in the medical dataset with a single voter. This is shown in Figure 6: on the left is the raw data, which to the right has been made *4-anonymous*[38].

However, without too much difficulty, it is clear to see that *k-anonymity* is vulnerable to two forms of attack.

A homogeneity attack would take advantage of one or more sensitive fields within *k* rows of a *k-anonymised* dataset being equal, allowing the sensitive value to be predicted for an individual. For example, in the second table in Figure 6, lifted from Machanavajjhala, A., Kifer, D., Gehrke, J., & Venkitasubramaniam, M. (2007), Alice and Bob are antagonistic neighbours and Bob falls ill. Alice finds only the second table but she knows that Bob is between 30 and 40 and his zip code, as he lives next door to her. She can then infer that

38 Machanavajjhala, A., Kifer, D., Gehrke, J., & Venkitasubramaniam, M. (2007)

Figure 6: Examples of anonymised datasets

	Non-Sensitive			Sensitive
	Zip Code	Age	Nationality	Condition
1	13053	28	Russian	Heart Disease
2	13068	29	American	Heart Disease
3	13068	21	Japanese	Viral Infection
4	13053	23	American	Viral Infection
5	14853	50	Indian	Cancer
6	14853	55	Russian	Heart Disease
7	14850	47	American	Viral Infection
8	14850	49	American	Viral Infection
9	13053	31	American	Cancer
10	13053	37	Indian	Cancer
11	13068	36	Japanese	Cancer
12	13068	35	American	Cancer

	Non-Sensitive			Sensitive
	Zip Code	Age	Nationality	Condition
1	130**	< 30	*	Heart Disease
2	130**	< 30	*	Heart Disease
3	130**	< 30	*	Viral Infection
4	130**	< 30	*	Viral Infection
5	1485*	≥ 40	*	Cancer
6	1485*	≥ 40	*	Heart Disease
7	1485*	≥ 40	*	Viral Infection
8	1485*	≥ 40	*	Viral Infection
9	130**	3*	*	Cancer
10	130**	3*	*	Cancer
11	130**	3*	*	Cancer
12	130**	3*	*	Cancer

Source: Machanavajjhala, A., Kifer, D., Gehrke, J., & Venkitasubramaniam, M. (2007)

he is represented by one of rows 9, 10, 11 or 12. All four of those people share the same sensitive column: cancer. She has deduced, therefore that Bob has cancer.

A background knowledge attack is less certain. Again, from Machanavajjhala, A., Kifer, D., Gehrke, J., & Venkitasubramaniam, M. (2007), Alice's friend Umeko is Japanese, in her 20s and lives within the zip code range beginning with 130. Therefore, Umeko must be represented by one of the individuals in the first four rows. There is no way of knowing whether Umeko has heart disease or a viral infection. However, with the background knowledge that Japanese people generally have an extremely low incidence of heart disease, Alice can conclude with some certainty that Umeko has a viral infection.

This potential attack can be thwarted using a method to maximise *l-diversity*, in which there must be more than *l* distinct values for the sensitive attribute. A table modified satisfies *l*-diversity if every cluster satisfied *l*-diversity[39]. This is known as *distinct l-diversity*. Variants include *entropy l-diversity*, in which the entropy of distribution of sensitive values in a cluster must be at least log(*l*) and *probabilistic l-diversity* in which the most frequent sensitive value in a cluster must be no greater than the reciprocal of *l*. Yet, there are issues with the *l-diversity* measure too and countless issues with successive fixes. No means of security is perfect; any GIS privacy policy must make use of these tools and measures and understand to what depth researchers and policymakers must go in anonymising data.

39 Song, F., Ma, T., Tian, Y., & Al-Rodhaan, M. (2019)

If policymakers and implementers cannot convince the public that their methods are secure and do not infringe on privacy, the benefits of GIS-based systems in improving service delivery will never be realised. We have explained the theoretical aspects of why GIS should be used and now we go on to answer our second research question with our work in Kerala: What are the on-ground steps and challenges that need to be overcome to build out an effective GIS system for the state of Kerala?

7. Building a GIS-based system: A case study of Kerala

The state of Kerala, located in Southern India, is an intriguing case in point. While the World Bank classifies India as a lower-middle income country, Kerala is already a largely middle-class, urban economy[40]. It contributes nearly 4% to national GDP despite accounting for only around 3% of India's total population. Moreover, it ranked first in the 2018 state-level Human Development Index, boasting the highest literacy rate and life expectancy in the country[41]. It is also India's least corrupt state, according to Transparency International India. Today, to further its growth trajectory, the Government of Kerala is adopting GIS-based systems, i.e. it is harnessing location-based data to streamline governance, and improve local planning and public service delivery.

The government turned to GIS after devastating floods engulfed the state in July and August of 2018. The floods were the worst the state had seen in nearly a century; one-sixth of the population lost lives, livelihoods and/or property. According to the government's calculations, the natural disaster cost the state around $2 billion. In the aftermath, policymakers realised the importance of location-based mapping, i.e., understanding the geographical extent of the state. Authorities lacked granular maps to pinpoint where citizens resided, where natural/artificial features lay, where the topography best facilitated the deployment of resources and so on. If knowledge of the spread of the state was available, policymakers could have activated faster and more informed responses to the floods.

40 Tandel et al. (2019)
41 Indiastat. (2018)

To rescue and recover from the event, the state's Chief Minister, Pinarayi Vijayan, established the Rebuild Kerala Development Programme (RKDP), terming it "a challenge and an opportunity to rebuild the State to ensure better standards of living to all sections of the society". Adopting a multi-sectoral approach, the state is associating with various partners who can provide analytical, advisory and implementation support for these projects. One of these organisations is IDFC Institute, a Mumbai-based think tank. We have partnered with the Rebuild Kerala Initiative (RKI), the state's coordination modality of RKDP, to assist various departments with their projects.

In particular, we have been invited to aid the technical and logistical implementation of different stages of the Mapathon Kerala project. Under the RKDP, the state government aims to build an open-data platform that will collect and make information about the state publicly available and accessible. Hence, Mapathon Kerala attempts to create a crowd-sourced localised map of the state which can be used to chart its geographical terrain, providing policymakers with key insights, such as for disaster planning (like in the case of the floods) and improving public service delivery on a day-to-day basis.

Mapathon Kerala is grounded in GIS technologies such as satellite- and drone-imagery, as well as the OpenStreetMap (OSM) platform[42]. OSM, much like Wikipedia, is open-source software that provides free access to add to and edit maps. Further, it emphasises local knowledge and community participation, allowing contributors to use aerial imagery, GPS devices and other software to update maps, maintaining data on railway lines, cafes, and so on[43]. The Mapathon Kerala initiative[44] is led by the Kerala State IT Mission, specifically by the Kerala State Spatial Data Infrastructure (KSDI), and is supported by the International Centre for Free and Open Source Software (ICFOSS) and the National Service Scheme.

Improving service delivery using GIS requires the Mapathon Project to be grounded in a robust technological base. The scope of our work includes building out the technical

42 Open Street Map (2020)
43 It is important to note that while the OSM layer developed for this project is specific to Kerala, GIS is not and can be replicated across states. States that create their own OSM layers can use them as a base for employing GIS.
44 Mapathon Kerala (2020)

infrastructure for the GIS-based system (in terms of data collection and database updation/management), and setting user interface guidelines for such tools to ensure interoperability of the data between government and service provider organisations. It becomes important to note that this work rests on safeguarding transparency and confidentiality, keeping citizens' rights to privacy in mind.

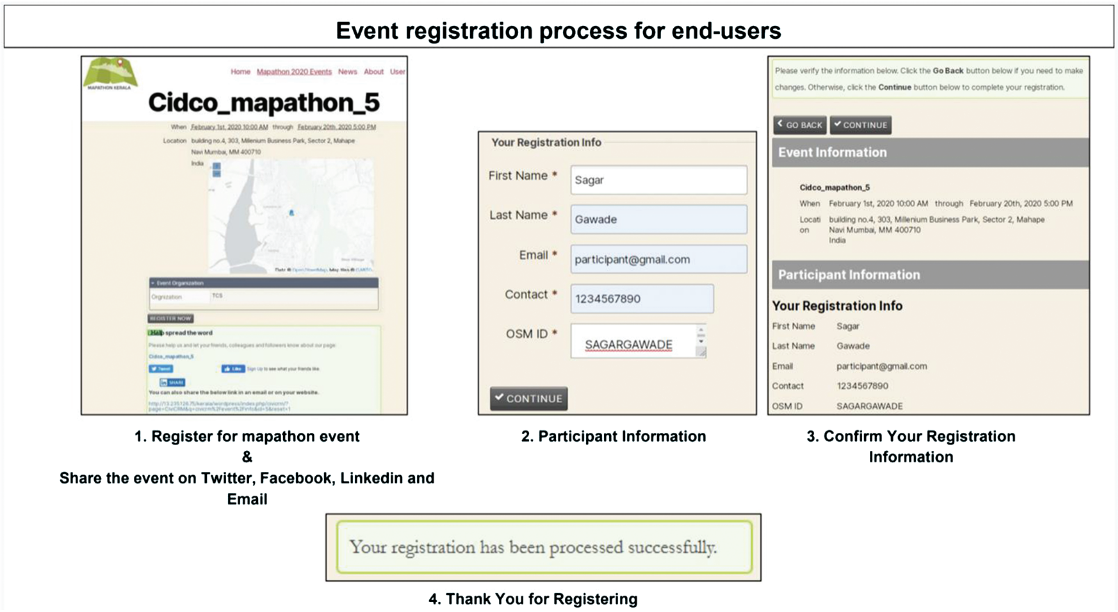
There are, fortunately, few privacy issues in the use of the Mapathon campaign and OSM. No personal data whatsoever will be recorded. Buildings, streets, schools and other public infrastructure clearly have no secrets in their location and function. Privacy issues may arise in users entering personal details either to use OSM itself or in publishing something they shouldn't. The former is up to the user; the latter falls under the same laws as publishing anything untoward. Our systems described below use only the user's OSM ID as a key between them; no extraneous data is stored or exposed.

As an aside, there have been issues with the concept of Mapathons in the past and we must be certain not to repeat those mistakes. Google launched its first Mapathon in 2013, inviting users to add to Google Maps. Government agencies lambasted the American company on various charges, primarily hinging on national security but extending to "polluting the web". One mistake Google made was simply being a foreign multinational which, politically, does not go down well in India; the other was in getting close to military areas in its mapping. The Kerala government will not suffer the first mistake and must be careful to avoid areas of national security.

In December 2019, we developed a campaign management platform using CiviCRM on an Amazon Web Services instance. CiviCRM is a web-based, open source Customer Relationship Management (CRM) software built by a community of open source contributors. To ensure usage and management of this system, services such as training sessions and email/social media campaigns are all going to be administered on this platform.

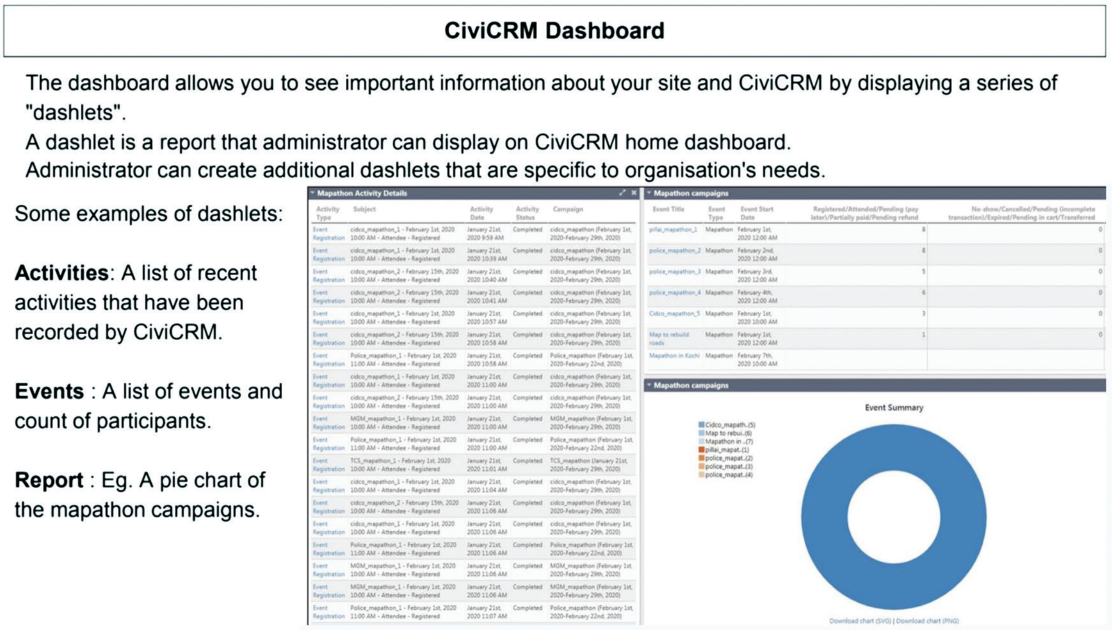
Additionally, via an API, we linked this server to a custom-built analytics system, which will directly interact with OSM. This is based on OSM Analytics. It can record

Figure 7: CiviCRM interface for Mapathon Kerala



Source: IDFC Institute

Figure 8: CiviCRM dashboard for Mapathon Kerala



Source: IDFC Institute

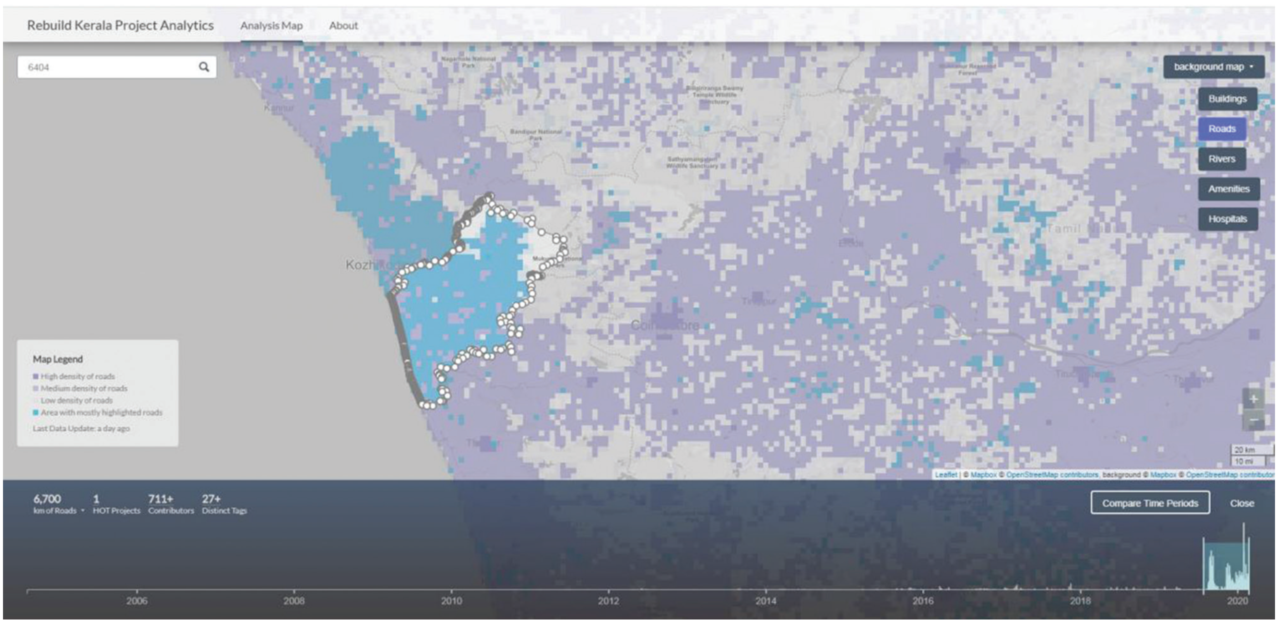
a list of recent activities, such as tabulating events or counting participants, as well as graphically document activities, such as the number of Mapathon Kerala campaigns.

This sort of information will be invaluable for those managing the initiative, giving them basic analytics not only on how the campaign is broadly running, but also allowing them to identify those who are performing well (and not so well) and so incentivising them further.

As a next step, we will build out this interface with analytics tools that can be employed to implement the project at a more granular level, i.e., organising mapping campaigns, tracking their work, setting up trainings, recording mapping activities and sharing their results with government departments in lieu of privacy guidelines.

Mapathon Kerala also fits into the RKI's long-term aim of citizen engagement with its Malayali communities. Through training sessions, this project is teaching thousands of volunteers to use OSM and its Android application to identify natural resources and man-made infrastructure, and upload these details onto the OSM platform. Via demonstration, volunteers are also learning how to use OSM in the Indian context. Some of these volunteers are being taught further to become trainers themselves to carry forward the programme. Anyone with a mobile phone and Internet connection can participate in this project. In addition to these government initiatives, we are also creating training modules and videos to illustrate the uses and applicability of GIS data (thereby also attempting to embed sustainability into the project), and developing

Figure 9: Snapshot of the OSM analytics platform showing the spatial visualisation of data mapped. This example shows which roads have been mapped using OSM.



Source: IDFC Institute

Figure 10: Government of Kerala's process for using GIS for the Mapathon Project

Source: Government of Kerala

gamification techniques of such infrastructure to enhance community participation. These processes will help support and institutionalise local level capabilities and scale the initiative across the state.

Through our work so far, we have gained several insights into the nuances and on-ground challenges that accompany deploying a technology (in particular, a GIS-based system) for good governance. Some of the lessons we have learned are outlined below.

7.1. Embrace technology and build expertise

As technological development is inevitable, states will need to catch up to further their growth. We have seen Kerala's policymakers readily embracing technology and data. Moreover, it is not just government agencies, even teachers of geography in the state employ arguably excess technology in their classrooms. Further, instead of aiming for complete accuracy, the government has accounted for a sound margin of error, which will, on net, yield several benefits to citizens. However, this willingness necessarily needs to be accompanied by technical knowledge, the lack of which can be a major hindrance in adopting new technologies. Hence it is important to augment technological literacy, particularly in local languages. We are converting all our material into Malayalam to facilitate programme implementation.

7.2. Institutionalise processes and monitor outcomes

For new initiatives to be successful, they need to be institutionalised in the government machinery. However, so far, we have observed that the training sessions conducted by the government for the Mapathon Kerala project have largely depended on a handful of people with the required technical skills. Unless the entire team is able to learn and document the process for undertaking these trainings, they will continue being one-off occurrences without the assurance of sustainability. Moreover, the number of trainings organised and the number of people trained are not always tracked, making follow ups difficult. Effective monitoring of progress is essential for the initiative to be scaled appropriately.

7.3. Encourage stakeholder engagement

Policy initiatives require effective cooperation with all stakeholders relevant to the programme. For Mapathon Kerala, we've seen how the community has stepped up to contribute to the government's undertaking. However, more training modules need to be conducted, especially to inform groups with different interests, so that the government can cover various aspects via a holistic approach. For example, volunteer groups invested in biodiversity conservation can be involved in mapping natural areas and other biodiversity indicators. Building these maps can also include an element of sensitisation around these topics, which can be used to raise awareness of environment conservation, natural heritage and disaster management. These outcomes can only be achieved with consistency in running training modules and facilitating data collection across the state. However, this requires political buy-in across the board to ensure consensus. A comprehensive engagement plan is critical in moving the project forward.

7.4. Make the software simple and usable

This is a key part of getting people involved, and is especially vital in the Mapathon project, which relies on crowdsourcing. We've seen how complex systems often fail, with initiatives going under-utilised. In this case, the government is working hard to make their forums attractive and so the analytics platform must aid in this, rather than just be a technical tool.

7.5. Coordinate among departments

With government departments functioning in silos, there is often information asymmetry. We've seen how inaccessibility of information and low government coordination can lead to knowledge remaining with only a few people, thereby limiting innovation and causing numerous open data problems. Hence, instituting easy and well-defined data sharing procedures between departments is essential.

GIS is extremely useful in improving public service delivery, particularly in healthcare, which was analysed in previous sections. As we move forward in this project, we will focus efforts on deploying the GIS-based system to improve such delivery in Kerala. The technology would be immensely advantageous in healthcare reform in the state. According to the World Bank and NITI Aayog's report, *Healthy States, Progressive India*, Kerala tops the country in best practices for healthcare. Both public and private healthcare systems are flourishing. The results are evident: the state has the highest life expectancy (as mentioned earlier) and the lowest infant and maternal mortality rates in the country; in terms of infrastructure, it has the highest number of medical institutions and hospital beds per capita across India. Further, a 2010 white paper published by the Economist Intelligence Unit lauded the Government of Kerala for its community healthcare model, particularly in its efforts to provide palliative care services. In fact, Kerala is considered to be healthier than most states in the U.S. (Rosling, 2009).

These successes can be attributed to the host of schemes the state government has initiated to improve its public healthcare system. Most notably, the Aardram Mission aims to create a "people friendly health delivery system" with a need-based approach. This scheme will transform PHCs into Family Healthcare Centres (FHCs), reaching a wider citizen group. Similarly, the Karunya Health Scheme was set up to provide health insurance coverage to low-income sections of society, ensuring that every citizen has fair and easy access to healthcare services. Moreover, through a universal healthcare policy, the state government funds several community-based programmes such as the 'Mother and Child Tracking System' and the 'Geriatric Care Project', thereby extending public services to all.

With these achievements under its belt, Kerala is well placed to adopt capacity building technologies such as GIS to catalyse its mission of increasing availability, accessibility and affordability of healthcare for its community.

First, GIS can be applied to identify healthcare trends in the state, and formulate solutions depending on the results. For instance, with demographic information mapped out, the data can reveal whether clusters of certain illnesses, such as diabetes or cancer, exist. Policymakers can then observe these trends, and respond to them accordingly, such as by deploying more medical personnel to these areas. A useful example of monitoring illnesses is that of Canada and the Nile virus mentioned earlier.

Second, going beyond merely tracking diseases, policymakers can use GIS technologies to understand where illnesses are likely to spread and manage outbreaks. Subsequently, they can systematically tackle these pockets by proactively implementing preventive measures. For instance, in 2014, a measles outbreak erupted in Disneyland in California. Authorities were able to use GIS to understand where infected children were staying and thus handle the spread of the disease. With GIS, the state government could institute better measures to contain viruses and diseases.

Third, GIS can streamline the delivery of health services. For instance, in the previously mentioned Bangladesh example, GIS was used to map out PHCs, understand access points and reallocate resources. More broadly, GIS can be used to create regional treatment plans and address wider healthcare needs. Thus, with targeted, data-driven strategies, authorities can enhance its healthcare system.

While applying GIS to healthcare, it is important to keep in mind the five components of an effective GIS mentioned earlier: spatially-referenced datasets, hardware tools, software, data sharing and management procedures and qualified people. It is often debated whether the location of a service should be determined by distance (i.e., distance travelled by a citizen for the service) or volume (number of citizens the facility serves). This question came up during the debates for Samagra Shiksha, an integrated school education policy in India[45]: Policymakers had the choice between distance, (placing the school at the centre where it is equally accessible by all students) and volume (a location closest to the largest number of students and providing buses for other students). The legislature ended up choosing distance over volume due to the poor quality of road networks, which would make it infeasible to run a large bus network. This serves as a useful example for policymakers to think about where to place new healthcare facilities. For instance, should PHCs be equidistant to the population served or should they be placed where the volume of those being served is highest, accompanied by ambulance networks? It is crucial for Kerala's policymakers to deliberate such issues when using GIS to improve healthcare delivery in the state.

Thus, while Kerala's adoption of GIS is laudable, sustaining it will be critical in improving service delivery. To scale its first venture, the Mapathon project, the government will need to focus on building expertise, institutionalising processes and monitoring outcomes, engaging with stakeholders, keeping the software simple and enhancing the information flow between agencies. Beyond Mapathon Kerala, these will be important to keep in mind in healthcare and other sectors for any state looking to employ GIS to improve service delivery. GIS, when adopted systematically, can create maximum impact with limited staff and facilitate better coordination across departments that ultimately benefits communities.

8. Policy Recommendations & Conclusions

Given the above, we recommend that the Kerala government:

1. Build up a GIS using the methods of Aghajani, et al. (2017):
    a. Spatially-referenced datasets
    b. Hardware tools
    c. Software

45 Ministry of Human Resource Development and Department of School Education and Literacy (2019)

d. Data sharing

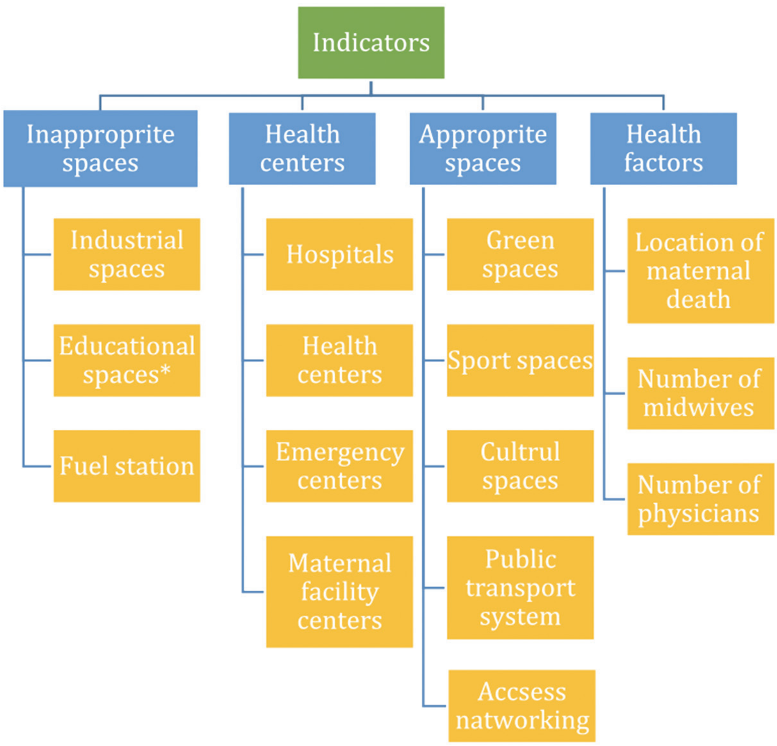
e. Qualified people

2. Consider the indicators and layers necessary in order to improve service delivery. This could follow the model put in place by Salehi and Ahmadian (2017) on maternal healthcare, drawn out in Figure 11.

3. Use appropriate techniques to mask data, especially given that elements will be on open networks. This must consider possible triangulation attacks that take advantage of completely independent, undeterminable data sets. This would include everything from deciding which data to supress to using jittering techniques.

4. Implement serious security protocols for all data that could even potentially be used to identify individuals. This must include encryption of data and hashing of passwords as well as strict rules on who has access to the most sensitive data and how they obtain that access. This must extend to how the data is transmitted across networks.

Figure 11: Indicators and layers to use GIS for healthcare



Source: Salehi, F., & Ahmadian, L. (2017)

5. Set up an interdepartmental group in order to ensure political buy-in from all stakeholders. All must understand the benefits of engaging with the project and how that offsets any costs they may incur. There must be communication between technical and non-technical staff, eased by training modules.
6. Different datasets are collected at different frequencies making it difficult to compare them on a single GIS platform. Hence, the increase in speed of collection of healthcare data (as recommended by the NDHB) should be carried out to allow for temporal analysis of different datasets.
7. Set up an open, public-facing platform (including OSM) with privacy safeguards. Doing so can provide the foundation for creating and giving public access to useful data services. This, in turn, would increase public engagement through transparency and promote innovation by start-ups.
8. Quantify (develop a delta) as to whether decision making is seeing improvement following the implementation of a GIS-based system.

These actionable policy recommendations are based upon our theoretical analysis of why GIS should be used by policymakers as well as on the on-ground lessons we have learnt while building a GIS-based system in Kerala. While our guidelines only touch upon healthcare, they could easily be extrapolated (and customised, where appropriate and necessary) for other sectors from policing to education. Ultimately, these recommendations can help provide a useful roadmap and toolkit for Kerala or any state government that is interested in employing GIS-based systems to improve public service delivery without infringing on the privacy of its citizens.

References

Aghajani, J., Farnia, P., & Velayati, A. A. (2017). Impact of geographical information system on public health sciences. Biomedical and Biotechnology Research Journal (BBRJ), 1(2), 94.

Armstrong, M. P., Rushton, G., & Zimmerman, D. L. (1999). Geographically masking health data to preserve confidentiality. *Statistics in medicine*, *18*(5), 497-525.

Ballany, S., & Nair, B. (2002). Application of Satellite Imagery and GIS in the preparation of development plans: a case study for Tirupati region. Indian Cartographer, 245-253.

Census Bureau, US Government (2020). Protection of Confidential Information. https://www.census.gov/about/policies/privacy/data_stewardship/title_13_protection_of_confidential_information.html

Chandra, V. (2016). Bharatmaps - Multi-layer GIS Platform Under Digital India. National Informatics Centre. Retrieved December 11, 2019, from https://negd.gov.in/sites/default/files/Multi%20Layer%20GIS%20Platform%20in%20Digital%20India_Vishnu%20Chandra_0.pdf

Chen, A. (2020, January 29). Why it's time to rethink the laws that keep our health data private, The Verge. https://www.theverge.com/2019/1/29/18197541/health-data-privacy-hipaa-policy-business-science

Chiemeka, E. C., Chukwudebe, G., & Ekwonwune, E. N. (2019). Development of a GIS Mapping Application for Crime Analysis for Law Enforcement Agencies in Owerri Metropolis. Development, 6(4).

Curtis, A. J., Mills, J. W., & Leitner, M. (2006). Spatial confidentiality and GIS: re-engineering mortality locations from published maps about Hurricane Katrina. *International Journal of Health Geographics*, *5*(1), 44.

Devulapalli, S., & Padmanabhan, V. (2019, June 19). India's police force among the world's weakest. Retrieved February 26, 2020, from https://www.livemint.com/news/india/india-s-police-force-among-the-world-s-weakest-1560925355383.html

Elwood, S., & Leszczynski, A. (2011). Privacy, reconsidered: New representations, data practices, and the geoweb. *Geoforum*, *42*(1), 6-15.

Forge, S. (2018, April). Optimal Scope for Free-Flow of Non-Personal Data in Europe. Retrieved February 26, 2020, from https://www.europarl.europa.eu/RegData/etudes/BRIE/2018/618988/IPOL_BRI(2018)618988_EN.pdf

Garfinkel, S., Abowd, J. M., & Martindale, C. (2019). Understanding database reconstruction attacks on public data. *Communications of the ACM*, *62*(3), 46-53.

Garg, A. (2016, May). Legal Challenges to Mapping in India #1 - Laws, Policies, and Cases. Retrieved from https://cis-india.org/openness/legal-challenges-to-mapping-in-india-1-laws-policies-cases#2

Geographic Information Systems (GIS) Use Cases, American Red Cross. https://www.preparecenter.org/sites/default/files/gis_final_use_cases.pdf

Gosselin P, Lebel G, Rivest S, et al. The Integrated System for Public Health Monitoring of West Nile Virus (ISPHM-WNV): a real-time GIS for surveillance and decision-making. Int J Health Geogr 2005;4:21.

Grant, A., Razdan, R., Shang, T. (2017). Coordinates for Change: How GIS Technology and Geospatial Analytics Can Improve City Services, Government Designed for New Times, McKinsey. https://www.mckinsey.com/~/media/McKinsey/dotcom/client_service/Public%20Sector/GDNT/GDNT_Coordinates_for_change_GIS_FINAL.ashx

Hao, K. (2020, March 24) Coronavirus is forcing a trade-off between privacy and public health. https://www.technologyreview.com/2020/03/24/950361/coronavirus-is-forcing-a-trade-off-between-privacy-and-public-health/ Citing 危机时刻的公共利益与个人隐私 https://mp.weixin.qq.com/s/2KRGP2ErKlQ9XF98asy8-w

Health & Human Services, US Government (2020). Health Information Privacy. https://www.hhs.gov/hipaa/index.html

Healthy States Progressive India. (2019). World Bank and NITI Aayog. http://social.niti.gov.in/uploads/sample/health_index_report.pdf

Human Development Index. (2018). Indiastat. https://www.indiastat.com/economy-data/8/human-development-index/14992/stats.aspx

India Corruption Survey 2018. (2018). Local Circles and Transparency International India. http://transparencyindia.org/wp-content/uploads/2019/04/India-Corruption-Survey-2018.pdf

Keola, S., M.Tokunaga, K.T. Nitin and W.Wisa, 2002.Spatial Surveillance of Epidemiological Disease A Case Study in Ayutthaya Province, Thailand, GIS development Magazine.

Kolios, S., Vorobev, A.V., Vorobeva, G.R., Stylios, C. (2017). Geographic Information Systems. Springer.

Krishnan, V. (2019, July 5). Report Summary: Economic Survey 2018-19. Retrieved February 26, 2020, from https://prsindia.org/sites/default/files/parliament_or_policy_pdfs/Economic Survey 2018-19 Summary-To Upload.pdf

Lawal, O., & Anyiam, F. E. (2019). Modelling geographic accessibility to primary health care facilities: Combining open data and geospatial analysis. Geo-Spatial Information Science, 22(3), 174-184.

MEITY (2019). Office Memorandum of constitution of Committee of experts to deliberate on data governance framework. https://meity.gov.in/writereaddata/files/constitution_of_committee_of_experts_to_deliberate_on_data_governance_framework.pdf

Machanavajjhala, A., Kifer, D., Gehrke, J., & Venkitasubramaniam, M. (2007). l-diversity: Privacy beyond k-anonymity. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, *1*(1), 3-es.

Mapathon Kerala (2020). https://mapathonkeralam.in.

Ministry of Health and Family Welfare (2019). Final Report On National Digital Health Blueprint. https://mohfw.gov.in/sites/default/files/Final%20NDHB%20report_0.pdf

Mesgari, M. S., & Masoomi, Z. (2008). GIS applications in public health as a decision making support system and it's limitation in Iran. World Applied Sciences Journal, 3(1), 73-77.

Ministry of Human Resource Development and Department of School Education and Literacy (2019), Samagra Shiksha, http://samagra.mhrd.gov.in/docs/Framework_IISE%20_F.pdf

Open Street Map (2020). https://www.openstreetmap.org/.

The Personal Data Protection Bill, 2019. Retrieved February 26, 2020, from https://prsindia.org/billtrack/personal-data-protection-bill-2019

Prakash, P. (2016, June). CIS's Comments on the Draft Geospatial Information Regulation Bill, 2016. Retrieved April 20, 2020, from https://cis-india.org/internet-governance/blog/comments-draft-geospatial-information-regulation-bill-2016

PRS Legislative Research (2016). The Geospatial Information Regulation Bill, 2016. https://www.prsindia.org/uploads/media/draft/Draft%20Geospatial%20Bill,%202016.pdf

The Quality of Death: Ranking End-of-life Care Across the World. (2010). The Economist Intelligence Unit. http://graphics.eiu.com/upload/QOD_main_final_edition_Jul12_toprint.pdf

Robin, T. A., Khan, M. A., Kabir, N., Rahaman, S. T., Karim, A., Mannan, I. I., ... & Rashid, I. (2019). Using spatial analysis and GIS to improve planning and resource allocation in a rural district of Bangladesh. BMJ Global Health, 4(Suppl 5), e000832.

Rosling, H. (2009). Asia's Rise-How and When. TEDIndia. https://www.ted.com/talks/hans_rosling_asia_s_rise_how_and_when/transcript?language=en

Ross, C. (2020, April 8). After 9/11, we gave up privacy for security. Will we make the same trade-off after Covid-19?. https://www.statnews.com/2020/04/08/coronavirus-will-we-give-up-privacy-for-security/

Salehi, F., & Ahmadian, L. (2017). The application of geographic information systems (GIS) in identifying the priority areas for maternal care and services. *BMC health services research*, *17*(1), 482.

Samarati, P., & Sweeney, L. (1998, June). Generalizing data to provide anonymity when disclosing information. In *PODS* (Vol. 98, No. 10.1145, pp. 275487-275508).

Schwab, K. (2017). The fourth industrial revolution. Currency.

Singh, A., Raghavan, M., Chugh, B., & Prasad, S. (2019, October 5). The Contours of Public Policy for Non-Personal Data Flows in India. Retrieved February 26, 2020, from https://www.medianama.com/2019/09/223-non-personal-data-flows-policy/

Song, F., Ma, T., Tian, Y., & Al-Rodhaan, M. (2019). A New Method of Privacy Protection: Random k-Anonymous. *IEEE Access*, *7*, 75434-75445.

Srivas, A. (2017, March). How the Controversial Geospatial Bill Snowballed – And Was Then Shoved Into Cold Storage. Retrieved from https://thewire.in/government/controversial-geospatial-bill-snowballed-shoved-cold-storage

Survey of India (2005). National Map Policy- 2005. http://surveyofindia.gov.in/files/nmp/National%20Map%20Policy.pdf

Sweeney, L. (2002). k-anonymity: A model for protecting privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, *10*(05), 557-570.

Tandel, V., Hiranandani, K., & Kapoor, M. (2019). What's in a definition? A study on the suitability of the current urban definition in India through its employment guarantee programme. Journal of Asian Economics, 60, 69-84.

Thorbecke, C. (2020, February 11). China launches coronavirus tracking app as death toll surpasses 1,000. https://abcnews.go.com/Business/china-launches-app-combat-coronavirus-spread/story?id=68907706

Uber Technologies, Inc. (2020). Available at: https://movement.uber.com

The UK's geospatial data infrastructure: challenges and opportunities. (2018, November). Retrieved February 26, 2020, from http://theodi.org/wp-content/uploads/2018/11/2018-11-ODI-Geospatial-data-infrastructure-paper.pdf

UN Sustainable Development Goals 2030. https://sustainabledevelopment.un.org/?menu=1300

Venugopala, Rao K., Ramesh, B., Bhavani, SVL., & Kamini, J. (2010). Urban and Regional Planning. Remote Sensing Applications: A book by National Remote Sensing Centre, 109-133.

World Bank Data for India. https://data.worldbank.org/?locations=IN-XN

Xinhua (2020, February 10). China introduces novel coronavirus close contact detection app. http://www.xinhuanet.com/english/2020-02/10/c_138770415.htm

Yang, Y., Zhu, J. (2020, February 7). Coronavirus brings China's surveillance state out of the shadows, Reuters. https://www.reuters.com/article/us-china-health-surveillance/coronavirus-brings-chinas-surveillance-state-out-of-the-shadows-idUSKBN2011HO

Zandbergen, P. A. (2014). Ensuring confidentiality of geocoded health data: Assessing geographic masking strategies for individual-level data. *Advances in medicine*, *2014*.

Acknowledgments

The authors would like to thank Dr. Niranjan Rajadhyaksha, Venkatesh Hariharan, Prakhar Misra, Isalyne Gennaro and the anonymous peer reviewers for valuable feedback and comments. We are grateful to Vikram Sinha and Archana Nirmal Kumar for editorial inputs. We would also like to thank the participants at the two research seminars of the Data Governance Network where this paper was discussed. In particular, we would like to thank Gautham Ravichander and Thejesh GN for serving as discussants for the paper at these seminars. We are indebted to M. Sivasankar, IT Secretary, Government of Kerala, and his team, as well as the Rockefeller Foundation, without whom this work would not have been possible.

About the Authors

Harsh Vardhan Pachisia is an Associate at IDFC Institute. His research focuses on projects related to urban governance, data empowerment and privacy. Prior to joining IDFC Institute, he worked as an Economics Teaching Fellow at the NYU Stern School of Business. Harsh has also been a Research Assistant to several NYU professors analysing topics such as 'Shadow Banking around the World' and 'Social Media Entrepreneurship'.

Girish Gupta is Head of Data Science at IDFC Institute. Previously, he was a foreign correspondent based in Venezuela where he produced ground-breaking investigations, covered daily clashes and reported on an unprecedented economic, social and humanitarian crisis. He wrote widely-used web and mobile apps to provide live and historic data on the country's economy. He later expanded the software into a global platform, Data Drum. He also worked across Latin America and the Middle East.

Kadambari Shah is a Senior Associate at IDFC Institute. Her research focuses on ease of doing business in India, special governance zones, and urban governance. At St. Xavier's College, Mumbai, she was on the editorial board of several scholarly journals and has presented research papers on development issues ranging from the middle-income trap to the role of race and gender in electoral politics.

datagovernance.org dgn@idfcinstitute.org

@datagovnetwork /datagovnetwork /datagovnetwork