DATA DIAGNOSTIC FOR KERALA

Spotlight on Resilience

Action plan based on a rapid diagnostic of data governance in the State of Kerala

THE WORLD BANK
IBRD • IDA | WORLD BANK GROUP

GFDRR
Global Facility for Disaster Reduction and Recovery

HUMAN RIGHTS INCLUSION EMPOWERMENT

The World Bank
1818 H Street NW
Washington DC 20433
Telephone: 202-473-1000
Internet: www.worldbank.org

Table of Contents

Acknowledgments

This publication is a product of close and constructive collaboration between the World Bank and the Government of Kerala with funding from the Global Facility for Disaster Reduction and Recovery (GFDRR). GFDRR is a multi-donor partnership that supports low- and middle-income countries to understand, manage, and reduce their risks from natural hazards and climate change. The publication was also supported by the Human Rights, Inclusion and Empowerment Umbrella Trust Fund as part of the "Mainstreaming a Rights-Based Approach to Information and Data in Bank Operations" program.

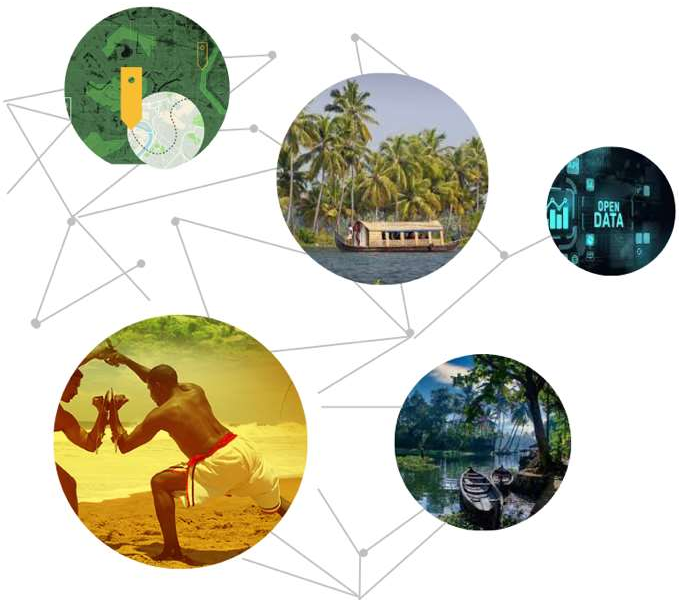
Several people provided inputs and contributed to the report. Abhas Jha provided overall guidance. The preparation of the report was led by Ravi Kumar. Illika Sahu, Prasanna Lal Das, Ravi Kumar and Shashank Jayakumar are co-authors. The authors gratefully acknowledge contributions from Nicholas Jones and Yim Chew Kloe Ng. Emily Kallaur provided editorial services. Natsuko Kikutake, Deepak Singh and Elif Ayhan provided valuable feedback. The team owes debt of gratitude to Pierre Chrzanowski, Thomas Danielewitz, Stephen Davenport, and Avni Sinha for serving as peer reviewers and providing detailed comments. The team is grateful to have worked closely with officials from the Government of Kerala comprising of over 18 departments, including Dr. Rathan Kelkar, Principal Secretary of Electronics and IT Department; K. Mohammed Y Safirulla, Deputy CEO of the Rebuild Kerala Initiative; Dr. Sekhar Kuriakose, Member Secretary of the Kerala State Disaster Management Authority; and other World Bank colleagues based in Washington, DC and New Delhi.

The diagnostic team also wishes to thank the wide range of stakeholders who readily agreed to be interviewed for the study, and whose input and feedback were essential to this report. The complete list of interviewees is included in the appendix.

Disclaimer

This diagnostic is not a comprehensive or representative assessment of the data ecosystem in Kerala, India as a whole, or for the area of focus (resilience). It does not provide an overview of every kind of potential data or open data program that could be developed. The diagnostic is not based on detailed, legal due diligence and does not constitute legal advice. Accordingly, this report should not be taken as a definitive statement on the topics it covers, but should rather be used as an indicative guide for further evaluation of the enabling policy, legal, and regulatory framework (including institutional aspects) for data in Kerala, and for further exploration of potential data use cases and opportunities.

This work is a product of the staff of the World Bank with external contributions. The findings, interpretations, and conclusions expressed in this report do not necessarily reflect the views of the directors or executive directors of the World Bank Group or the governments they represent. The World Bank Group does not guarantee the accuracy of the data included in this work.



List of abbreviations

AI	Artificial Intelligence
API	Application Programming Interface
AR	Augmented Reality
CDO	Chief Data Officer
CSO	Civil Society Organization
DCRA	Dynamic Complete Risk Analysis
EU	European Union
GDS	Government Digital Service
GoK	Government of Kerala
HoD	Head of Department
IKM	Information Kerala Mission
IoT	Internet of Things
IT	Information Technology
JPEG	Joint Photographic Experts Group
KDSAP	Kerala Data Sharing and Accessing Policy
KILA	Kerala Institute of Local Administration
KSDI	Kerala State Spatial Data Infrastructure
KSDMA	Kerala State Disaster Management Authority
KSDP	Kerala State Data Policy
KSITM	Kerala State Information Technology Mission
KSREC	Kerala State Remote Sensing and Environment Centre
KWRIS	Kerala Water Resource Information System
LIMS	Land Information Management System
LSGD	Local Self-Government Department
MNREGA	Mahatma Gandhi National Rural Employment Guarantee Act
NDEM	National Database for Emergency Management
NDMA	National Disaster Management Authority
NDRF	National Disaster Response Force
NIC	National Informatics Centre
OSM	Open Street Map
PDF	Portable Document Format
PforR	Program for Results
PNG	Portable Network Graphics
PWD	Public Works Department
ReLIS	Revenue Land Information System
RKI	Rebuild Kerala Initiative
VR	Virtual Reality
WB	World Bank
WDR	World Development Report
WRD	Water Resources Department

Executive summary

Motivated by lessons learned from the devastating 2018 floods, the Government of Kerala's (GoK) Rebuild Kerala Initiative (RKI) provides a design blueprint - embedding principles of sustainability, equity, inclusiveness, and effectiveness - for a new way of thinking about development in the State of Kerala. The GoK has invested significantly in data, recognizing it as a critical resource with the potential to help deliver more effective and efficient state services to residents and businesses. Improving the collection, sharing, and utilization of data constitutes a crucial component of the blueprint. Better data governance can make societies more resilient through helping governments and companies anticipate and reduce risk, improve disaster preparedness, respond more effectively to events, and recover from disasters more efficiently and equitably. A powerful example comes from Odisha, where strategic investments in data-driven warning, communication, and response tools following the 1999 cyclone in which almost 10,000 people died, led to a dramatic reduction in lives lost in a subsequent storm (fewer than 40 people) and a reduction in economic damage.[1] Kerala itself offers another example – the quality and extent of its disease surveillance activities were an important factor in its successful response to the COVID-19 pandemic.[2]

The GoK is committed to using data-driven tools and services for resilience, especially after the 2018 flood event. It has embarked upon several innovative data programs that address known gaps in the state's resilience-related service delivery architecture and data ecosystem. The World Bank (WB), through the ongoing Resilient Kerala Program for Results (PforR, P174778) Project and Additional Financing (P177980) Project, is supporting the GoK in its efforts to use data effectively for resilience. The objective of the PforR is to enhance Kerala's resilience against the impacts of climate change and natural disasters, including disease outbreaks and pandemics. The WB aims to support the GoK in combining data initiatives related to climate change and disaster risk management into an integrated ecosystem of technology products and processes, as well as strengthen institutional mandates by enhancing data governance policies and creating incentives for data sharing.

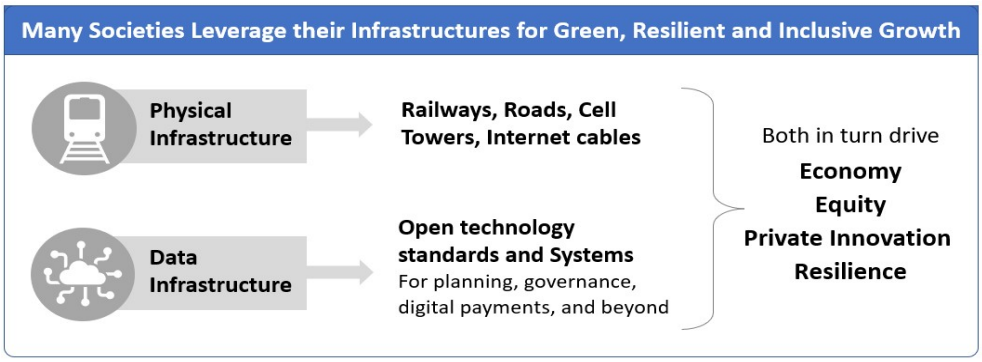
The envisioned shared digital ecosystem would support more timely and effective decision making around climate and disaster risk (Figure 1). This will enhance availability of reliable information, enhance interaction between sectors to encourage innovation, and build institutional memory to support longevity of technology initiatives. The concepts of the global open data movement can be applied to reduce vulnerability to natural hazards and climate change, increase public access to risk-related information, communicate risk more effectively to decision makers, and give policymakers access to new forms of data-driven risk assessments, thus increasing the state's capacity to manage and respond to crises.

At the invitation of the RKI, a WB team conducted a rapid diagnostic of the state of data governance in Kerala, with a special emphasis on programs and initiatives linked to the state's resilience agenda. The diagnostic was carried out through secondary research and semi-structured interviews (see Appendix 2 for a list of interviewees) based on the eight pillars[3] of the Open Data Readiness Assessment (ODRA) methodology[4] developed by the WB. The Bank has deployed the ODRA in about 30 countries over the last decade to help governments assess their readiness to establish, manage, and benefit from open data programs. For the Kerala diagnostic, the team adapted the toolkit, increasing the focus on resilience use

[1] See https://www.wunderground.com/cat6/Damage-Widespread-Fanis-Death-Toll-Remarkably-Low-Thus-Far.
[2] See https://gh.bmj.com/content/5/7/e003212.
[3] The pillars are leadership, policy/legal framework, institutional structures/capabilities, government data management, funding, demand for data, civic engagement, and technology and skills infrastructure. Please see Appendix 3 for more details.
[4] Available at http://opendatatoolkit.worldbank.org/en/odra.html.

cases and concrete items for action, and broadening its scope beyond open data to align it better with GoK priorities and relevant WB operations.

Figure 1. Leveraging Physical and Data Infrastructure for Development



This rapid diagnostic is not designed to be a comprehensive assessment but rather to serve as a key input into the State of Kerala's ambition for an open data initiative to support resilience (as laid out as one of the priority areas in the GoK's Rebuild Kerala Development Plan). Through the diagnostic process and the World Bank partnership established with the state following the 2018 floods, it is clear there is an abundance of pertinent data being collected and key resilience datasets are being identified and cleaned into usable formats. However, these are often not maintained or shared between departments in a systematic manner. Therefore, this diagnostic aims to better understand Kerala's current data sharing environment and recommend how a sustainable system for data sharing can be institutionalized. To further the GoK's ambitions and actions on data sharing for resilience, the ongoing PforR Project will support implementation of the recommendations of this diagnostic, specifically through open data targets included in the Additional Financing (P177980) to the PforR. These targets include conducting an open data readiness diagnostic, creating a roadmap for leveraging open data to strengthen climate and disaster resilience, piloting open data initiatives in climate sensitive sectors, and training public and private sector stakeholders on open data for building climate and disaster resilience.

The initial findings of the rapid diagnostic suggest that a vibrant, innovative, and entrepreneurial data ecosystem for resilience exists within the GoK. The government and its partner agencies have developed and deployed several sophisticated resilience-related, data-driven tools, applications, and platforms such as a geo-portal, an open data website, primary data collection tools (Gramam[5], PWD4U[6]), and platforms (eSevanam[7]). The government is also in the early stages of developing fundamental applications like a Unified Registry (a reference application that proposes to provide a unified data architecture for the state) that incorporates data architecture and design ideas consistent with similar applications in countries with mature data environments. The main findings section includes a table that lists key GoK applications and initiatives relevant to the resilience agenda and provides a preliminary heuristic analysis.

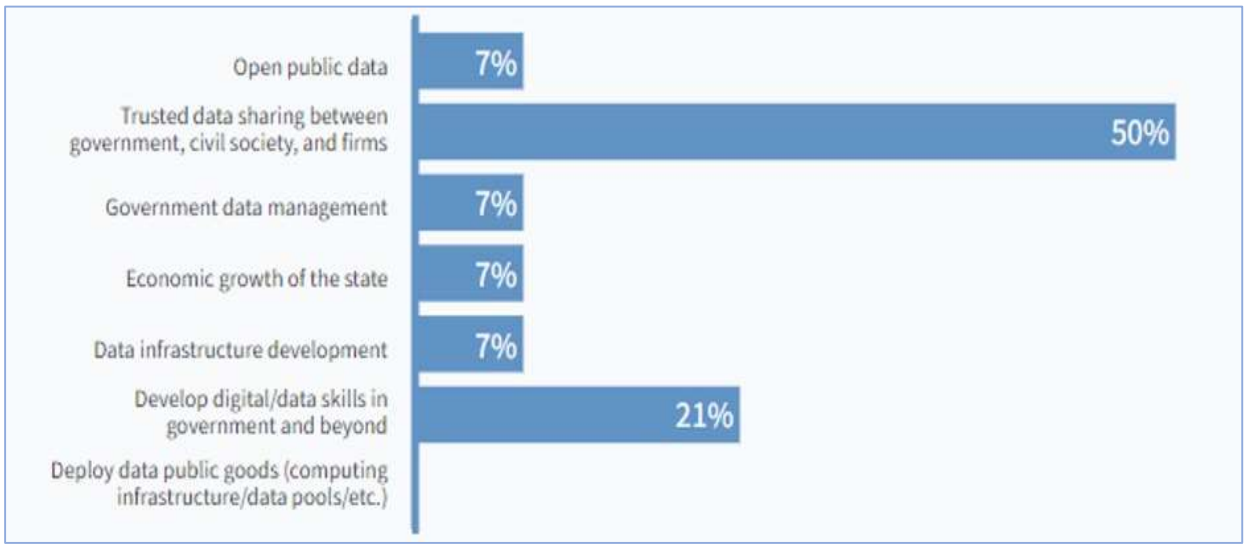
However, the GoK is not yet reaping all the potential benefits of its data-related investments to date, as most of these initiatives are not underpinned by a common set of standards, methods, and policies. In May 2023, the Chief Minister of Kerala Pinarayi Vijayan declared Kerala as India's first fully e-governed state, digitizing government services for prompt and transparent service delivery to citizens and

[5] A mapping application by the GoK that relies on local volunteers to supplement mapping data.
[6] "PWD 4U is part of the digital initiative by the Government of Kerala to enable citizens to notify the authorities of any defects or problems in the road assets managed by Kerala PWD." The public app is available on Google Play.
[7] "eSevanam is the official service portal of the Government of Kerala. See https://services.kerala.gov.in/.

empowering its people by bridging the digital divide.[8] However, the government has not yet clearly articulated the idea of a central, cross-cutting data authority (as distinct from IT or e-governance functions) across the GoK. The lack of a clear data governance framework undermines the effectiveness of these digital initiatives and leads to suboptimal citizen user experiences. Part of the challenge is at the leadership and strategic level, where the top leaders' vocal commitment to data-driven resilience has not yet been matched by codified institutional structures and processes. Instead, ad hoc arrangements and informal practices dominate. Complicating matters further, legislative work is still ongoing at the national level in India on core issues like data privacy and protection. The state data policy, under development, reflects some of the challenges created by the gaps in national data policy.

From a data governance perspective, there is scope for more strategic management of many government applications for resilience. Many of these applications appear to have been conceptualized and developed without effective engagement with non-government actors and citizens, and often without proper coordination with other government agencies that may be developing similar services or collecting/sharing related data. The GoK does utilize several potentially effective engagement channels, such as mapathons/hackathons, community training programs, and entrepreneurship platforms, but these are not always used strategically or with adequate follow-up (see Figure 2 for a snapshot of what some stakeholders think should be the state's data priorities). This leads to a lack of use cases that can visibly demonstrate the value of the initiatives and increase the utilization of data and related tools. The supply-driven nature of many initiatives results in ineffective metrics to track the performance of the programs. Many government data and digital initiatives also struggle to attract and retain people with the right skills to implement and scale services, which creates knowledge gaps and wobbly execution.

Figure 2. Informal poll of government partners about what the state's data priorities should be



Source: World Bank staff, based on informal poll conducted during a 3-hour workshop in Kerala.

Appendix 1 and 3 provide a more detailed description of the main findings of the rapid data diagnostic, along with the methodology used to collect information for this report.

[8] See https://www.livemint.com/news/india/kerala-becomes-indias-first-e-governed-state-cm-pinarayi-vijayan-announces-11685032093902.html

Action steps

This report recommends a set of strategic and tactical action steps that the GoK can take to strengthen data governance and demonstrate the value of data-driven initiatives to achieve development outcomes.[9] The action steps are based on the rapid diagnostic and reflect the national experiences of countries such as Australia, Brazil, Canada, China, Finland, Japan, New Zealand, South Korea, the UK, the US, and others, which can serve as useful benchmarks for the GoK. The action steps also draw on national and state-level experience from India, which is increasingly recognized as an emerging model with respect to several data governance pillars.

STRATEGIC ACTION STEPS

Based on the findings of this diagnostic, it is recommended that the GoK consider the following key strategic steps to reinforce the foundations of the state's data ecosystem. These strategic actions address key government priorities and reflect leading global practices. Each action is described in more detail later in the report, alongside the Kerala context and examples from leading practices in different countries.

1. **Develop a state data strategy.** Data is recognized as a core component of the RKI. But there is no single document that articulates a data vision for the state or a comprehensive and systematic data strategy. Building resilience is a challenge that cuts across sectors and requires coherence and coordination within the GoK. The lack of an overarching data strategy leads to the misaligned and uncoordinated development of data tools and services by different agencies. As of the time of writing, more than 60 national governments have established data strategies, as have regional blocs like the European Union.[10] Interestingly – and possibly tellingly – only 3 percent of low-income countries have national data strategies, while 52 percent of high-income countries do.[11] The GoK should establish a multi-stakeholder working group to develop a state data strategy that will:

- Formally articulate the state's vision for data-driven development.
- Identify and prioritize initiatives and provide guidance for resource allocation and sustained investment.
- Incorporate a model for the state's data architecture that can provide direction on harmonizing existing and new data products and services in the state.
- Formalize the role of relevant data institutions and leadership positions.
- Serve as the vehicle for community engagement.
- Delineate the distinction between related concepts like a data-driven public sector, digital transformation, IT, and e-governance and establish data as a clearly defined functional area within government.
- Incorporate the GoK's approach towards emerging data-driven technologies such as artificial intelligence (AI), Generative AI, augmented reality (AR)/virtual reality (VR), distributed ledgers, and others.

A data strategy is distinct from a data policy. Policy generally refers to guidelines and rules that govern behavior and decision-making, while strategy is a unified, structured, and integrated plan of action or roadmap that is designed to achieve specific objectives of an entity. It should be noted that the GoK has

[9] The action steps described do not cover the entire spectrum of data governance activities that the state should consider. The diagnostic team focused largely on GoK entities and the report does not represent the perspective of businesses, civil society organizations, academia, and regular citizens who are essential participants in developing trust in data/digital systems. The report also does not consider infrastructure and cybersecurity – either within the government or in the environment of citizens and end users – both of which can have a significant impact on the success of data initiatives.
[10] https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy_en.
[11] World Bank. 2021. "Data for Better Lives." World Development Report. https://www.worldbank.org/en/publication/wdr2021.

already drafted data policy documents, and the World Bank team has provided feedback as mentioned in this report.

2. **Establish clear leadership and institutional mandates on data, supported by adequate financing.** There are currently differing views about where data leadership rests in the state (as distinct from leadership on IT or e-governance, for example), and which specific agencies and personnel within them serve as nodal actors. Countries such as the US and UK offer relevant organizational models (as discussed later in the report) that the government may consider. The GoK should:

- Establish the position of Chief Data Officer (CDO) for the state, who would report directly to the Chief Minister or the Minister of Finance. The position should be formalized through a legal provision that also defines the responsibilities and mandate of the CDO and their office.
- Officially identify, through the same legal provision, a nodal agency for data governance in the state. The Kerala State IT Mission (KSITM) is the nodal IT agency for the state, but its data mandate is inconsistently interpreted by different stakeholders in the absence of a clear directive from the government. Some of the confusion is the result of an unclear distinction between IT and data within the GoK machinery.
- Identify/appoint nodal data officers in government agencies with clearly defined roles and responsibilities.

3. **Broaden the scope of, and finalize, the state data policy documents that are currently undergoing consultation.** Establishing the ambit of state data policy is tricky, especially when it comes to topics such as data privacy and protection for which national policies are still being finalized. While many foundational elements of data policy rest with the national government, the GoK should do the following, with an eye on the resilience agenda:

- Finalize state policy guidance on open data and spatial data, given their relevance to the resilience agenda; the government has already circulated draft documents for consultation.
- Emphasize the human aspects of data governance and expand the scope of data policy consultations to explicitly incorporate elements of a "social contract for data," including greater civic and business engagement, capacity development inside and outside government, and data innovation and entrepreneurship.
- Focus on data sharing and reuse. Develop clear policy guidance on data sharing among government agencies and clarify data collection responsibilities (based on the principles established in the state data strategy).

4. **Develop a plan to manage data-related staffing in the state government.** Developing the right skill base within the government is a major challenge for the GoK, which finds it difficult both to recruit and retain data talent (especially as the government is unable to offer market-competitive salaries). Reskilling the existing government workforce is a related task. The report considers the experience of countries such as Canada, Singapore, and others and suggests that the GoK should explore the following options, which have shown promise elsewhere:

- Develop staffing models to utilize external talent in government data initiatives by offering specific, time-bound government positions to external talent. Suggested options to explore include data fellowships, co-development programs with the private sector/academia, partnerships with CSOs that can serve as conduits for talent, and a "free agent" model for rapid, on-demand hiring.

- Expand career opportunities for data talent within the government. Suggestions include creating leadership development pathways within government, and talent marketplaces that encourage government staff mobility.
- Retrain public sector staff in data/digital competencies and invest in continuous upskilling.
- Establish a data/digital center for excellence within the GoK with both public and private participation.
- Continue to engage external firms (as vendors or suppliers) to develop major data systems (it is impractical to develop full stack data competency within the GoK).

5. **Grow the demand side of the data ecosystem within the state.** The GoK's data engagement model is inward looking, with relatively little attention paid to the demand for data, or to coordination with non-governmental actors around increasing data use and co-creation of products and services that communities need. This usually leads to the development of supply-oriented tools and services, rather than what is most beneficial to residents. Within the government, little attention is paid to metrics on demand-driven data publishing and sharing, and there is thus limited validation of what initiatives are most effective. The challenge is not unique to Kerala and the report offers examples drawn from the Netherlands, Finland, Serbia, and elsewhere of tools used to develop successful engagement models. The GoK should:

- Publish data strategically, with a purpose, and for reuse.
- Measure the demand for data.
- Identify and work with data intermediaries with strong local credibility and grassroots presence.
- Widen and systematize the range of engagement channels to reach different types of audiences.
- Explore co-development models to create products and services.

In addition to taking the strategic steps listed above, the state should consider the following themes. At this stage, discussions would be exploratory in nature, but could suggest ways to strengthen the data ecosystem (the proposed data strategy could be a useful mechanism to launch consultations):

1. **Explore a "data space" for resilience in the state.** The European Union has recently introduced the concept of "data spaces," meaning "data infrastructure with tailored governance mechanisms that… enable secure and cross-border access to key datasets in the targeted thematic area." Data spaces cut across organizational boundaries and engage a wide variety of government and non-government stakeholders that share data, infrastructure, and solutions. The EU's Green Deal data space, which offers features such as an AI workbench, a data marketplace, and a community toolkit, may be a relevant example. The field of resilience, which is a multidisciplinary and cross-sectoral theme, can offer a practical testing ground to explore the concept of data spaces in Kerala.

2. **Incorporate more use of machine data in service design and delivery.** Internet of Things (IoT) devices have become integral to data environments, including and especially the resilience infrastructure. Many state agencies already deploy sensors to gather data, but their use is inefficient and uncoordinated. To broaden the use of IoT within the state, the GoK should (based on experience in countries such as Japan, Germany, and others):

- Establish a sandbox program to test the use of machines to develop/deliver resilience applications and services.
- Establish linkages with other governments/agencies with experience incorporating machine data in their operational processes.
- Develop capacity programs within government to grow awareness of IoT among government officials.

TACTICAL ACTION STEPS

The diagnostic team has identified several tactical steps that the GoK should consider (the suggested steps are a mixture of action items that are already underway but largely uncoordinated, and actions to fill gaps observed during the diagnostic). These include activities such as developing an inventory of data collection tools, expanding/retiring tools based on gaps/duplication, formally tracking demand for GoK data, both from within the government and from outside and establishing related metrics, and initiating the process of identifying and documenting the ownership of the "golden" source of each data variable that the GoK collects and uses across agencies. The Tactical Steps section of the report provides the complete list of proposed tactical action items.

Use cases

The report provides examples drawn from resilience-related data use cases that have been successfully deployed in different parts of the world. It also includes a brief analysis of the main lessons learned, which include the importance of enabling data flows across organizational boundaries, the need to develop use cases from the perspective of citizens, the significance of experimentation and continuous learning, the imperative to place a premium on a user-oriented approach, and the realization that successful use cases combine offline and online elements.

The report cautions that while there is great urgency and enthusiasm within the GoK to deploy new use cases that demonstrate the power and value of data-driven tools, it is difficult to successfully implement and scale use cases without an adequate data governance framework to support them. The GoK should consider new use cases in association with the strategic and tactical steps described earlier in the report. A few techniques related to developing use cases that it may consider include:

- Design from the outside in and engage stakeholders – especially external ones – at the outset.
- Think beyond government data.
- Establish and/or leverage data innovation hubs that bring a range of stakeholders together, including those who work on data and those who work on other aspects of resilience.
- Engage the appropriate institutions.
- Focus on people, not technologies.

More detail about each suggestion and finding is provided later in the report, alongside a description of the findings of a workshop organized by the WB mission team in Thiruvananthapuram. The report will be of particular interest to GoK policymakers, civil society leaders, academics, and the private sector. The report would also be of high interest to development partners, including World Bank staff, and donors.

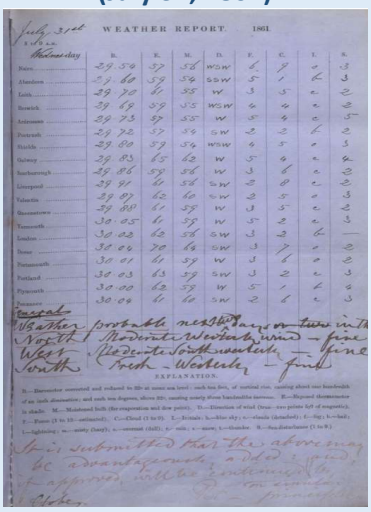
The report is organized into distinct sections for clarity and readability. First, background information and the methodology for the diagnostic are presented. Next, the main findings, along with recommended strategic and tactical action steps, are highlighted to emphasize key takeaways. Finally, use cases and appendices supplement the report with further details and examples.

Background

Context: Data and resilience

For centuries, data has been used to understand weather or disease patterns to protect people and property. The UK's Meteorological Office was established in 1854 to safeguard sailors and ships by analyzing weather patterns in the Atlantic Ocean (Figure 3). Weather data, including on rain, thunder, and sea disturbances, was collected and transmitted via the telegraph system from the British and Irish coasts to the Meteorological Office in London at 9am every day.[12] In 1859, a devastating storm struck the British Isles, causing the deaths of over 800 people and destroying more than 130 ships. This disaster made the need for a national storm warning system clear. To prevent future losses, the Meteorological Office began further analyzing the data it had been collecting since 1854 and established a Storm Warning service. The first storm warning was issued in 1861, but many lives were still lost in areas where the warning was ignored.[13] Furthermore, in 1858, London was struck by a cholera outbreak, and Dr. John Snow, a physician, was able to identify the areas with the highest concentration of cases by mapping the related deaths alongside the city's water pumps. This led to the removal of the contaminated water pump and almost entirely stopped new cases from occurring.[14]

Figure 3. The world's first public weather forecast (July 31, 1861)



Source: The Met Office, UK Government

The World Bank's 2021 World Development Report (WDR) *Data for Better Lives* puts forward a vision for achieving development objectives by harnessing the power of the 21st century data revolution. The amount of data continues to increase exponentially around the world, and if used creatively, data offers previously unimaginable ways to increase productivity and improve socioeconomic outcomes. Investments in data transparency can also provide measurable economic benefits.[15] The WDR argues that leveraging data effectively depends on establishing a new "social contract for data" that builds trust in the data ecosystem and ensures that benefits are shared equitably. Overall, it posits that "data can improve social and economic outcomes, but only if they are used systematically in ways that create information that generates insights that improve lives." This requires that governments take a holistic approach to data use that prioritizes strengthening institutions, data analytics capacity, stakeholder engagement, and the broader enabling environment as part of a coherent strategy, rather than pursuing one-off or short-term investments in single datasets or stand-alone data systems. In this context, the purpose of this diagnostic is to help the GoK advance its vision of "a resilient, green, inclusive and vibrant Nava Keralam" by more effectively collecting, sharing, and utilizing data for more effective, efficient, and resilient state services for residents and businesses.

Data and information have become critical tools to generate value and to address development challenges, with the world at the threshold of what has been called the Fourth Industrial Revolution.

[12] See https://www.metoffice.gov.uk/research/library-and-archive/archive-hidden-treasures/robert-fitzroy.
[13] See https://www.metoffice.gov.uk/research/library-and-archive/archive-hidden-treasures/royal-charter.
[14] World Bank. 2021. "Data for Better Lives." World Development Report. Washington, DC: World Bank. Available at https://www.worldbank.org/en/publication/wdr2021.
[15] See https://openknowledge.worldbank.org/entities/publication/5dad280a-b4a8-538b-a34b-407980c8a848

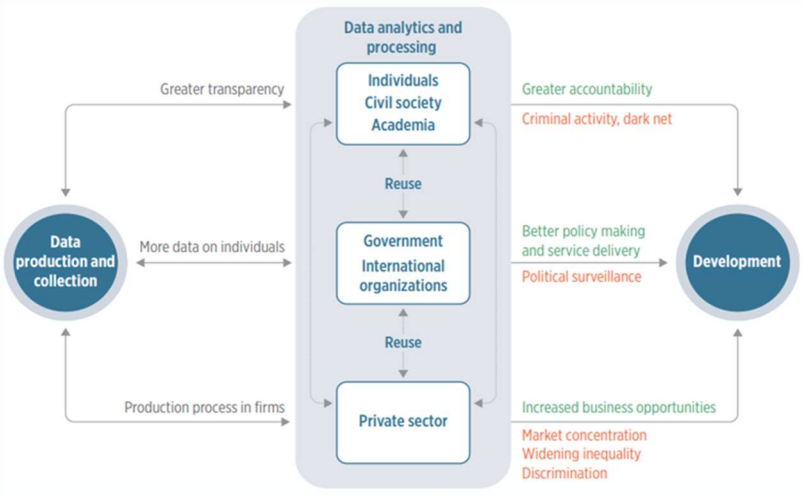
The global increase in data production and digital connectivity, and exponential advances in data analytics capabilities, have coalesced to become a powerful engine for change in both positive and negative ways. To evolve accordingly, governments need to fill development data gaps and turn data into actionable information. This means establishing or strengthening key institutions to scale up data-driven decision-making and support improved outcomes for the poor and marginalized, taking a rights-based approach to the provision of information. Moreover, "for data to maximize value, the data should have adequate coverage (be complete, frequent, and timely), be of high quality (be accurate, comparable, and granular), be easy to use (be accessible, understandable, and interoperable), and be safe to use (be impartial, confidential, and appropriate)."[16]

There are various definitions of data, as this term has evolved in the past few decades. The 2021 WDR acknowledges this lack of uniformity regarding the concept of data, and aligns with Carriere-Swallow and Haksar´s definition: "data can be quantitative or qualitative in nature, and may be stored on analog (that is, paper, stone tablets) or digital media". Furthermore, the 2021 WDR establishes the differences of data and information, as "data must be processed, structured, and analyzed to be converted into information".

Data can be divided, depending on its main purpose, into public and private intent data. Public intent data refers to data that is primarily generated by the public sector, although Civil Society Organizations (CSOs), academia and international organizations can also contribute to this process. The WDR classifies public intent data in six categories: (a) administrative data; (b) census; (c) sample surveys; (d) citizen-generated data; (e) machine-generated data; and (f) geospatial data. This data aims to improve decision-making by the state and improve the relationship between the state and society. These kinds of data can improve lives if they lead to the creation of better targeted service delivery and policies, and prioritization of scarce resources, particularly to help marginalized populations. In situations where official data does not exist on neglected or underrepresented populations, citizens or civil society groups may be able to shine light by generating data and demanding action or accountability.

In contrast, private intent data refers to "data collected and curated by the private sector for commercial purposes." Innovations in the use and application of data by businesses are creating tremendous economic value by enhancing data-driven decision-making and reducing transaction costs. While the GoK might benefit from implementing solutions around both kinds of data, public intent data and the role of government in the data ecosystem are the focus of this diagnostic.

Figure 4. How data can support development: A theory of change



Source: World Bank WDR, 2021. Note: Positive impacts show in green, negative impacts in red.

Data has significant potential to foster development, if used responsibly. Governments, civil society and the private sector increasingly initiate and manage data initiatives to

[16] See: https://openknowledge.worldbank.org/entities/publication/d00af143-c4da-5b0c-99f5-7e6f55a8fe2d

accomplish their objectives. Such initiatives can concern collection, production, use, and/or dissemination of data, and can support the government at the sub-national and national levels in designing, implementing and monitoring policies and programs. According to the WDR's theory of change, shown in Figure 4, leveraging data successfully can enable 1) greater accountability; 2) better policy making; and 3) greater business opportunities in the private sector.

Government decision making, including in times of disaster or crisis, requires accurate, detailed, representative, and timely data and information. Greater availability of data, and the ability to process it effectively, can be critical to improved policymaking and service delivery. Again, data can be generated by the government itself (i.e., census data), by academics and NGOs (data stemming from their research or contact with the citizenry), and even the private sector (data on traffic accidents by insurance companies). Ideally, to be useful for service delivery management, it should be possible to disaggregate data by different user groups based on geography, gender, income status, vulnerability, etc. This enables the government to understand who can access services (or not) and how their experiences differ. The data should also ideally be linked to actual service delivery events so that government officials can follow up with specific service providers to take corrective action. Data must also be representative of users if the intent is for government to have a reliable picture of the average service delivery experience. For instance, citizen feedback that requires the user to proactively submit a comment or complaint can be useful in identifying issues, but will not be representative of the average user experience. Data proactively collected by the government and data proactively submitted by citizens can complement each other to improve government's understanding of the user experience.

Globally, there are numerous cases where data has helped governments and communities strengthen their resilience. Box 1 provides an example from Japan. In Brazil, government authorities and academics collaborated to create the "Digital Disaster Atlas".[17] It is now helping the government better understand disaster risks in the country. In the Philippines, there is GeoRiskPH,[18] a portal that has data from "21 national government agencies, three nonprofit organizations, and more than fifty local governments". This helps citizens "imagine the impact of disasters" and plan, and enables policymakers to assess how many people and what kind of infrastructure might be impacted by a certain type of disaster such as a flood or tsunami.[19]

The 2014 Ebola outbreak in Sierra Leone was a major public health crisis. To face the challenge and minimize its impact, the government in collaboration with aid agencies and CSOs realized the importance of data sharing in a context where this kind of flow of information was not the norm. Shared data included information about cases, availability of medical facilities and healthcare workers, resources to fight the epidemic, relevant medical information for citizens, and so on. Furthermore, much of this information was geo-referenced and useful maps were developed. As a result, healthcare workers, citizens, policymakers and aid agencies were able to make informed decisions that contributed to more effective control of the epidemic.[20]

In Costa Rica, the government created the National Climate Change Metrics System.[21] This platform incorporates data from a wide range of relevant agencies. Through this collaboration, data generation, processing and analysis will contribute to more accurate information about climate change drivers, risks

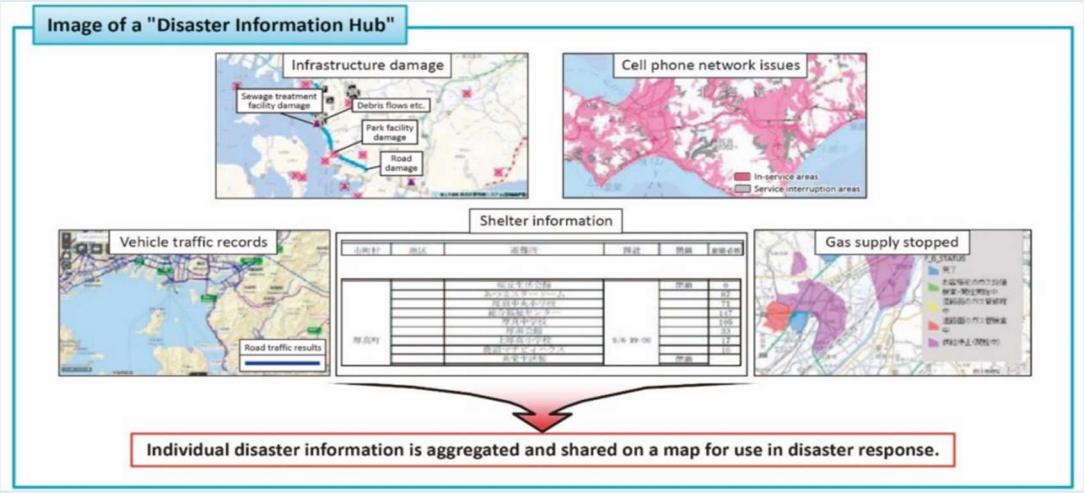
[17] Available at https://www.emdat.be/emdat_atlas/sub_html_pages/sub_html_BRA.html.
[18] Available at https://georisk.gov.ph/.
[19] See https://www.esri.com/about/newsroom/blog/philippines-shared-disaster-imagination-supports-resilience/.
[20] Andrew Young and Stefaan Verhulst. 2016. "Open Data Impact: When Demand and Supply Meet." GovLab and Omidyar Network: March 2016. Licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. Available at http://odimpact.org/files/open-data-impact-key-findings.pdf.
[21] National Climate Change Information System. SINAMMEC Costa Rica. Available at http://www.sinamecc.go.cr/.

and responses.[22] Ultimately, this will also enhance government reporting mechanisms under the Paris Agreement. Similarly, in Chile, government data on climate change has enabled CR2, a research institution, to build models that are now used by a wide range of stakeholders.[23]

Box 1. Snapshot: How Japan uses data and AI for disaster management

As weather patterns become more unpredictable and damage caused by typhoons and floods becomes even more severe, the Government of Japan is leveraging emerging digital technologies to detect and mitigate risks. Over the last decade, a series of strategic steps has been taken to strengthen the country's resilience. For example, Japanese officials are using sensors installed in electricity poles and waterways to detect increases in water levels and inform authorities so they can take necessary actions. Officials are also piloting chatbots to instantly communicate relevant data to citizens and encourage them to proactively act to protect their assets and lives.[a]

In 2016, after realizing that it was difficult to understand the "movements of evacuees and the status of relief supplies" in response to the Kumamoto Earthquakes, the government determined it was critical to develop "a system for a disaster information hub among relevant organizations in advance".[c] The primary goal of Japan's AI Technology Strategy of 2022 was to establish "systems and technological infrastructure to maximize protection of people's lives and property against pandemics and large-scale disasters."[b] Used well, AI systems could enable populations to get ready in advance of disasters. At the same time, Japan's experience offers the lesson that without reliable and timely data, AI is not very helpful; for example, AI is not able to predict earthquakes in advance because of the lack of data on which to base predictive models.[d]



Source: The Government of Japan.

[a] See https://www.japantimes.co.jp/news/2022/12/31/national/disaster-preparedness-tech/
[b] See https://www8.cao.go.jp/cstp/ai/aistratagy2022en_ov.pdf.
[c] See https://www.bousai.go.jp/en/documentation/white_paper/pdf/2021/R3_hakusho_english.pdf.
[d] UNESCO. 2021. "UNESCO Science Report: The race against time for smarter development." Available at https://www.unesco.org/reports/science/2021/en/download-report.

[22] Delfina Grinspan and Jesse Worker. 2021. "The Payoffs of Open Climate Data." World Resources Institute: March 2021. Available at https://medium.com/opendatacharter/the-payoffs-of-open-climate-data-c5c1016665ad.
[23] Ibid.

However, for data to bring about all these benefits, it is critical to mitigate risks that arise from the increased production and accessibility of sensitive information. As shown in Figure 4, alongside its positive effects, the data revolution has the potential to cause serious negative impacts. Policymakers and partners need to be mindful of the potential for unintended consequences as they design data-related interventions. Increased production and availability of individually-identifiable information, for example, creates opportunities for politically-motivated use of personal data, or surveillance by public or non-public entities. According to the WDR, if "public accountability is strong and state actors can be presumed to act in the broader public interest," these risks can be managed, but such conditions cannot be assumed. Cybercrime that exploits sensitive data, sometimes facilitated by the "dark net", has a staggering annual cost. The private sector can also potentially misuse and exploit personal data, and data-driven businesses often experience increasing returns to scale that increase market concentration.

Risks that accompany increased production and use of data underscore the need for a social contract for data, as outlined in the WDR. This needs to be enforced via a data governance framework that "can strengthen trust in the data system, thereby incentivizing the use of data-driven products and services, increasing their value, and ensuring a more equitable distribution of benefits".

Background and methodology: The Kerala data diagnostic

Kerala is highly vulnerable to natural disasters and the dynamics of a changing climate. Kerala is prone to natural hazards such as cyclones, monsoon storm surges, coastal erosion, sea level rise, tsunamis, floods, droughts, lightning, landslides, land subsidence and earthquakes. Kerala's State Disaster Management Plan identifies 39 types of known and reported hazard types that can turn disastrous without proper preparedness and risk reduction planning. With Cyclone Ockhi in 2017, floods and landslides in 2018, 2019, and 2020, and the COVID-19 pandemic, the state has experienced major disaster events for four consecutive years. Preparing for potential future disasters requires systemically strengthening the government's capability to carry out an integrated and coordinated set of policy, institutional and budgetary changes over time.

Following the devastating flood in 2018, the GoK committed itself to going "beyond traditional approaches to recovery and reconstruction to not only recover fully from the current disaster but also to prepare better for future disasters."[24] This vision led to the creation of the Rebuild Kerala Initiative. The Rebuild Kerala Initiative (RKI), which is housed in the Local Self Government Department (LSGD), has "a bold vision" and, among other goals, aims to use data to help "build the foundations for a resilient, green, inclusive and vibrant Nava Keralam". [25]

Kerala's experience during the COVID-19 pandemic serves as an example of data playing a critical role in helping a government and community respond to a crisis. The GoK was able to effectively manage and control COVID-19 in Kerala. On January 30, 2020, Kerala became the first state in India to be affected by COVID-19. Just a few weeks later, by early March, the state had the highest number of cases in the country. By building on the lessons learned from the Nipah epidemic in 2018-2019, Kerala government agencies collaborated to effectively screen and follow-up with every individual who entered the state, institute contract tracing, implement quarantining, engage the community, and mobilize residents for behavioral change. These government actions enabled the state to generate quality data on the status of the pandemic, slow down its spread, and effectively contain it.

[24] See https://www.preventionweb.net/files/67193_rebuildkeraladevelopmentprogramme.pdf
[25] World Bank Project Information Document (P174778), available at https://documents.worldbank.org/en/publication/documents-reports, and GoK framework for Rebuild Kerala Initiative. For more information see http://rki.lsgkerala.gov.in.

This diagnostic assessment was initiated as a response to a GoK request and undertaken in support of the ongoing "Resilient Kerala Program for Results", a World Bank project. A WB team conducted a rapid data diagnostic focusing on programs related to the state's resilience agenda, and on open data, which is enshrined as one of the cross-cutting pillars in the RKI framework.

The methodology for the diagnostic involved secondary research and semi-structured stakeholder interviews based on the eight pillars of the Open Data Readiness Assessment (ODRA) methodology[26] developed by the WB. The WB has deployed the ODRA in about 30 countries over the last decade to help governments assess their readiness to establish, manage, and benefit from open data programs. For the current diagnostic, the team adapted the ODRA toolkit by increasing the emphasis on concrete action steps and use cases, and broadening its scope beyond open data, to align it better with the priorities of the GoK and ongoing relevant WB operations.

The purpose of the ODRA methodological tool is to assist stakeholders in planning a government open data program at the national or sub-national level, or for an individual public agency. The eight pillars of the ODRA are: Senior Leadership; Policy/Legal Framework; Institutional Structures, Responsibilities and Capabilities within Government; Government Data Management Policies, Procedures and Data Availability; Demand for Data and Data Services; Civic Engagement and Capabilities for Data; Funding a Data Program; and Technology and Skills Infrastructure. Further details on the dimensions and corresponding questions posed can be found in Appendix 3. This diagnostic is intended to provide the GoK with concrete recommendations in support of its ambition to leverage data for development.

[26] The team adapted ODRA Version 3.1, prepared by the World Bank's Open Government Data Working Group; available at http://opendatatoolkit.worldbank.org/en/odra.html.

Main findings

The overall conclusions of the rapid diagnostic suggest that a vibrant, innovative, and entrepreneurial data ecosystem for resilience exists within the GoK. The GoK and its partner agencies have developed and deployed several sophisticated resilience-related, data-driven tools, applications, and platforms such as a geo-portal, an open data website, and primary data collection tools. The government is also in the early stages of developing fundamental applications like a Unified Registry that incorporates data architecture and design ideas consistent with similar applications in countries with mature data environments. The GoK is however not yet able to derive the full extent of benefits from these applications, as most of these initiatives are not underpinned by a common set of standards, methods, and policies, leading to suboptimal citizen user experiences and effectiveness. Main findings are organized by the eight pillars of the ODRA methodology. A short summary of main findings under each of the eight pillars is presented in Table 1, with a detailed account of the diagnostic report's findings in Appendix 1.

Table 1. Main diagnostic findings by Pillar

Pillar	Importance	Commentary
Senior Leadership	Very High	The leadership environment in Kerala is mixed – there is a strategic mandate for data-led development, there are strong individual digital leaders, and the state has developed several well-received digital applications, but the institutional leadership structure for data (versus IT) is still not fully developed and there is not yet any functional leadership for the data agenda specifically.
Policy/Legal Framework	Very High	While many foundational elements of data policy rest with the national government, the GoK has begun to formalize state-level data policy through two main data policy documents that are currently under consultation. The content of the documents suggests a growing focus on data sharing arrangements within government, standardization of data collection, and development of institutional arrangements for data governance within the state.
Institutional Structures, Responsibilities and Capabilities within Government	Medium High	There are several agencies and institutions in the GoK that have separate, sometimes overlapping mandates within the government's data ecosystem. Currently there is no agency with a clear mandate to play a leading role on data.
Government Data Management Policies, Procedures and Data Availability	High	Kerala is a highly digital government, and most government services are available online through a variety of digital channels. Some challenges were observed, such as the lack of a central data inventory, difficulty in acquiring data from national platforms, inconsistent data sharing practices within government, lack of comprehensive digitization, etc.
Demand for Data and Data Services	Very High	Even as the state has developed a number of digital platforms and services, very few agencies appear to work closely with non-governmental and private stakeholders to assess the demand for and the effectiveness of the solutions they have developed.
Civic Engagement and Capabilities for Data	High	The GoK mostly engages citizens through public communication portals where public participation is invited; citizen engagement has not yet extended to actively promoting the usage of open government data or involving the community in data gathering and maintenance of sensors/resilience monitoring.
Funding a Data Program	Medium High	Overall, agencies consulted for this diagnostic indicated they do not have reliable funding for their long-term data initiatives. Officials expressed their concern that there were no dedicated teams to work on funding schemes to sustain data initiatives over the long run.
Technology and Skills Infrastructure	High	The state government recognizes the importance of training and invests in trainings, study tours, and post-graduate courses on e-governance for GoK employees. However, more needs to be done as agencies cite lack of availability of relevant skills as one of their primary challenges.

Strategic action steps

This section describes recommended strategic action steps that would lay the foundation for effective data governance in the state of Kerala. The strategic recommendations reflect a combination of leading global practices and priorities expressed by GoK officials during the consultations for the diagnostic. The suggested steps, each discussed further below, are:

1. Develop a state data strategy.
2. Establish clear leadership and institutional mandates on data.
3. Finalize the state data policy documents currently under consultation.
4. Develop a plan to manage data/digital staffing in the GoK.
5. Grow the demand side of the data ecosystem within the state.

This section also includes a discussion around the following additional and exploratory strategic themes, which could broaden the ambit of the data strategy proposed above:

1. Create a "data space" for resilience in the state.
2. Incorporate more use of machine data in service design and delivery.

1. Develop a state data strategy

CONTEXT

The data ecosystem within the GoK is vibrant, innovative, and entrepreneurial, but it doesn't yet have a formally articulated strategic direction, resulting in a lack of clarity about leadership and prioritization of initiatives; duplication of effort; and inconsistent understanding of long-term goals and delivery mechanisms. There is also a lack of common understanding about the role of data within the GoK and how it overlaps with and is distinct from related concepts such as IT, digitizing the public sector, and e-government. While some in the state leadership believe strongly in the importance of data as a governance asset and a critical tool for building resilience, others have not fully internalized the idea.

LEADING PRACTICE

A lack of clarity around the overarching data agenda is not unique to Kerala, and has prompted governments around the world to develop national or local level data strategies. These strategies provide policy and regulatory guidance to maximize the social and economic value of data – personal, non-personal, public, and private. At the time of writing, more than 60 national governments have established data strategies, as have regional blocs like the European Union.[27] Interestingly – and possibly tellingly – only 3 percent of low-income countries have a national data strategy, while 52 percent of high-income countries do.[28] Many state and city governments have initiated their own data strategies as well;[29] prominent examples include cities such as London (UK)[30] and Buenos Aires (Argentina), and states/provinces such as Ontario (Canada)[31] and California (USA)[32]. The purpose of a local/regional data strategy is not to supplant or otherwise challenge the national data strategy – most local data strategies

[27] See https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy_en.
[28] World Bank. 2021. "Data for Better Lives." World Development Report. Washington, DC: World Bank. Available at https://www.worldbank.org/en/publication/wdr2021.
[29] See https://oecd-development-matters.org/2021/05/12/why-governing-data-is-key-for-the-future-of-cities/.
[30] See https://data.london.gov.uk/blog/data-for-london-a-city-data-strategy-for-a-smart-city-future/.
[31] See https://www.ontario.ca/page/building-digital-ontario.
[32] See https://innovation.ca.gov/pdf/CalData_Californias_Data_Strategy_2020.pdf.

cover the flow of data within and across the jurisdiction; focus on local issues, service delivery and implementation activities; and, where necessary, fill gaps in national strategies.

The goal of these strategies may differ depending on the context – in some jurisdictions they may be primarily concerned with competitiveness or economic growth; others may focus more on personal and social protections, and minimization of harm to vulnerable populations – but the operational aim is typically to make it easier to collect, access, share, use, and reuse data among people, groups, and organizations. A formal data strategy, even in the absence of accompanying legislation, can help either kickstart or provide direction to data initiatives in numerous ways:

- It is an important signaling tool that demonstrates leadership commitment to the agenda.
- It sets out a vision for data-led development, clarifies how better use of data can achieve broader strategic goals, identifies institutional arrangements, and establishes the broad principles of data governance in the jurisdiction.
- It, often implicitly, establishes a balance of power in the data ecosystem among the government, civil society, and the private sector.
- It can provide guidance on priorities for resource allocation and future investments.
- It describes a model for a data architecture that can guide harmonization of existing and new data products and services in the state.
- It establishes data governance as a separate functional area (distinct from related functions like digitizing the public sector, IT infrastructure, and e-governance).
- It elucidates the government's approach to emerging data-driven technologies such as AI, AR/VR, distributed ledgers, and others.

India does not yet have a formal national data strategy, though several government documents outline emerging data priorities. The draft National Data Governance Framework[33] policy sketches a vision of "maximizing data-led governance and catalyzing data-based innovation that can transform government services and their delivery to citizens". It is a relatively brief document but identifies many themes of interest and relevance to the GoK, including the idea of a whole-of-government approach, modernization of data collection and storage practices, platformization of data services, a new emphasis on non-personal data, and development of an institutional framework for data governance. A preceding discussion paper proposing an AI strategy for India[34] included a section on data ecosystems that covered several issues such as developing a large national corpus of data (which would be structured and/or annotated), improving data sharing frameworks (within government and with corporations), and greater community engagement with data and data crowdsourcing. India is updating its data-related legislation with a new data protection bill[35] under discussion (a new Digital India act is under development too).

KERALA

A state data strategy for Kerala could be an impactful way to address key challenges in the current state data ecosystem in the following ways:

- The GoK could initiate a discussion on, and potentially achieve agreement about, several fundamental data governance concepts and practices that are currently being considered in ad hoc fashion by different agencies and application developers. These include:
    - A whole-of-government approach to data. Sometimes also called "joined up government" or "one-stop government", the basic idea of "whole-of-government" is to move past agency-

[33] See https://www.meity.gov.in/writereaddata/files/National-Data-Governance-Framework-Policy.pdf.
[34] See https://indiaai.gov.in/documents/pdf/NationalStrategy-for-AI-Discussion-Paper.pdf.
[35] See https://www.meity.gov.in/data-protection-framework.

and department-level silos of data and offer "personalized and accessible public services, which are to be planned, implemented and evaluated with their participation, and the opportunities presented by the Internet to transform the way the government works for the people."[36]

- The autonomy principle. Drawn from medicine, the autonomy principle in data governance seeks to clarify that authorized agencies remain in control of their data governance practices even if their data is consumed by other agencies that may have different standards.
- The idea of government as a platform/digital public infrastructure. In some ways analogous to the whole-of-government idea, the concept of "government as a platform" typically covers the movement away from public silos. Instead, it offers the provision of reusable public digital tools and data by the government, co-production of policies and services, and a focus on effective, inclusive digital public services. Digital public infrastructure provides essential building blocks that can be leveraged or repurposed by third parties to provide citizen-facing services.
- The "Ask Once Only" principle. Popularized through Estonia's experience, the principle aims to ensure that citizens, institutions, and companies only must provide certain standard information to the authorities and administrations once. It is then incumbent upon public institutions to reuse the information instead of asking for it again.

- The GoK could clarify the continuum between open data and confidential data. In Kerala there is still a lack of clarity about what openness implies, and what data can and should be open to which actors. Sensitive data is not always cataloged as such. The process of preparing the strategy can provide a useful opportunity for a broad conversation about openness in a way that the routine policymaking process cannot.
- It would be an opportunity to formally define flagship data initiatives and identify, and potentially allocate, resources for them. The strategy can serve as a useful conduit for prioritization and uniform communication about the rationale behind flagship initiatives. Typically, such initiatives provide the data and policy infrastructure for other state applications.
- Through the strategy, the GoK could formally establish appropriate data institutions and their leadership, and articulate their mandate within the government. This may also include the formalization of a position of Chief Data Officer in government, as discussed below.

2. Establish clear leadership and institutional mandates on state data

CONTEXT

The data leadership picture in Kerala is unclear. The Kerala State IT Mission (KSITM) has the official mandate to lead on IT in the GoK and is deeply engaged in several core data initiatives in the state, including the proposed Unified Registry, the management of the state data center and the state's spatial data infrastructure, and the development of state data policy documents. It is however unclear that it has an unambiguous authorization to be the nodal data agency for the state. Consequently, other state agencies such as the Water Resource Department (WRD) and the Kerala State Disaster Management Authority (KSDMA) have established their own data labs, amassing the multi-sectoral data needed by their agency, while others such as the Kerala State Remote Sensing and Environment Center (KSREC), the Public Works Department (PWD), and others have developed many data applications without any formal

[36] UN. 2012. "United Nations e-Government Survey: Chapter 3, Taking a Whole-of-Government Approach". Available at https://publicadministration.un.org/egovkb/Portals/egovkb/Documents/un/2012-Survey/Chapter-3-Taking-a-whole-of-government-approach.pdf.

coordination with KSITM. Many of these apps are developed based on ad hoc needs, and sometimes they duplicate the features of existing applications used by similar actors. Similarly, the LSGD has its own data management apparatus and a suite of solutions (KSMART- most visibly) that do not necessarily follow any common state level data standards.

Part of the challenge is definitional or conceptual – the government has not clearly articulated the idea of a central, cross-cutting data function (as distinct from IT or e-governance functions), with formal leadership, across the GoK. The data strategy proposed in the section above, and accompanying government directives, would be important to establishing both. It is also vital not to rush into creating institutional arrangements and/or new roles in government without sufficient clarity about the role of data in the GoK and the necessary political and financial support for the agenda.

LEADING PRACTICE

The UK, following the success of its open data initiative and the establishment of the Government Digital Office (placed strategically within the Cabinet Office) in 2010 to make "digital government simpler, clearer and faster for everyone"[37], created the office of the Chief Data Officer (CDO) in 2015. This step aimed to address the challenge of "discrete projects" without "central coordination" that did not allow the "data agenda to move forward".[38] The key objectives of the UK's first CDO could be applied wholesale to the GoK:

- Overseeing the definition and enforcement of a new Government Data Standard.
- Championing open data, and opening up existing government data wherever possible.
- Driving the use of data as a tool for making decisions in government.

The United States government has followed a similar trajectory[39], establishing a foundational data strategy with the goal of "leveraging data as a strategic asset".[40] The government formally defined the office and responsibilities of the CDO through the Evidence-Based Policymaking Act[41], which requires all federal agencies to develop plans for "data the agency intends to collect, use, or acquire to facilitate the use of evidence in policy making" and appoint "an official with statistical expertise to advise on statistical policy, techniques, and procedures" for the task. The federal CDO council website provides a good job description[42] of the CDO role that can serve as the starting point for a similar role within the GoK. A fuller review of the US CDO position[43] comes from the consulting firm Deloitte, and envisions CDOs "playing an increasingly important part in delivering better public outcomes at the city, state, and national levels."

Many developing countries have begun to appoint CDOs as well. In 2021, Guangdong became the first Chinese province to create a CDO office in China.[44] Brazil similarly has established the office of a data governance head in government[45] after establishing its digital governance strategy in 2016 and creating a

[37] See https://www.gov.uk/government/organisations/government-digital-service.
[38] See https://www.techrepublic.com/article/uks-first-chief-data-officer-to-focus-on-making-data-a-public-asset/.
[39] The legal foundations include the Digital Accountability and Transparency Act of 2014 that required "the U.S. federal government to transform its spending information into open data" and provided a significant boost for open data. In 2018, the Geospatial Data Act became the law, as did the Evidence-Based Policymaking Act.
[40] Available at https://strategy.data.gov/.
[41] Available at https://www.congress.gov/bill/115th-congress/house-bill/4174.
[42] See https://www.cdo.gov/about-agency-cdos/.
[43] Deloitte Development. 2019. "The Chief Data Officer in Government: A CDO Playbook." Deloitte Center for Government Insights. Available at https://www2.deloitte.com/content/dam/insights/us/articles/5004_CDO-Data-strategy/DI_CDO-in-govt-playbook.pdf.
[44] See https://www.china-briefing.com/news/guangdong-pilot-program-on-chief-data-officer-accelerating-digitalization-and-achieving-better-data-protection/.
[45] See https://www.zdnet.com/article/brazilian-government-seeks-data-governance-head/.

central committee on data governance.[46] As in most other examples, these CDOs each lead a discrete entity with responsibilities that cut across multiple agencies and formal recognition as the implementer of the state's data mandate.

KERALA

As it deliberates on a potential CDO position, the GoK should consider the following:

- Position the office within the Chief Minister's Office or the Finance Ministry to visibly demonstrate the importance of the agenda and provide the fledgling entity with the necessary political support while it develops and demonstrates its value proposition (see the example of the UK government, where the role is placed within the Cabinet).
- Provide the necessary legal and regulatory support for the position in the form of legislation or government orders that are unambiguous and clearly define the responsibilities of the position/office and its relationship and working arrangements with other state agencies and ministries (see the legal framework in the US as an example).
- Develop an institutional structure to support the position in the GoK. The US federal government again provides a useful model that requires all government agencies to create their own data officer positions and through a CDO Council provides a useful governance apparatus.
- Staff the agency with appropriate technical expertise ("data experts not generalists" in the words of an interviewee), and provide it the necessary independence from political pressures (the CDO in the US, for instance, is a non-political appointee).
- Clearly distinguish the CDO role from that of the Chief Information Officer role, which tends to be more IT-focused. While they sound similar, it is important to clarify that the CDO is focused on delivering value from data assets in the state (within the government but also throughout the entire data ecosystem) and not on developing/maintaining/growing government IT systems.

3. Broaden the scope of and finalize the state data policy documents

CONTEXT

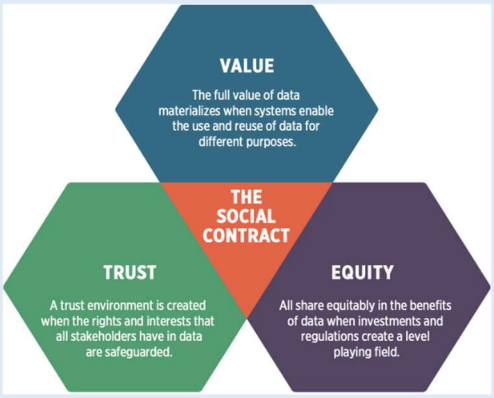
The GoK is currently considering drafts of two data policy documents – the Kerala State Data Policy (KSDP-2022), developed by the Department of Electronics and Information Technology, and the Kerala Data Sharing and Accessing Policy (KDSAP – version 3.0), developed by KSITM. The two documents complement the Draft Kerala Spatial Data Sharing Policy (KSDSAP 2020), developed by KSREC. Each of these documents draws on its counterpart issued by the national government. The policies reflect ongoing attempts by the state administration to bring order and clarity to its data policy environment, which is frequently marked by ad hoc and informal arrangements that rely on personal relationships and initiatives rather than systematic tools. The content of the policy documents also suggests a growing focus on data sharing arrangements within government, standardization of data collection, and development of institutional arrangements for statewide data governance.

LEADING PRACTICE

The World Bank's 2021 World Development Report, *Data for Better Lives,* provides a useful framework to review the data policy initiatives in Kerala. The report envisages a "Social Contract" for Data that has three equally important pillars (Figure 5):

[46] See https://www.tandfonline.com/doi/full/10.1080/25741292.2022.2065065.

Figure 5. World Development Report 2021 Social Contract Framework



Source: World Bank WDR, 2021.

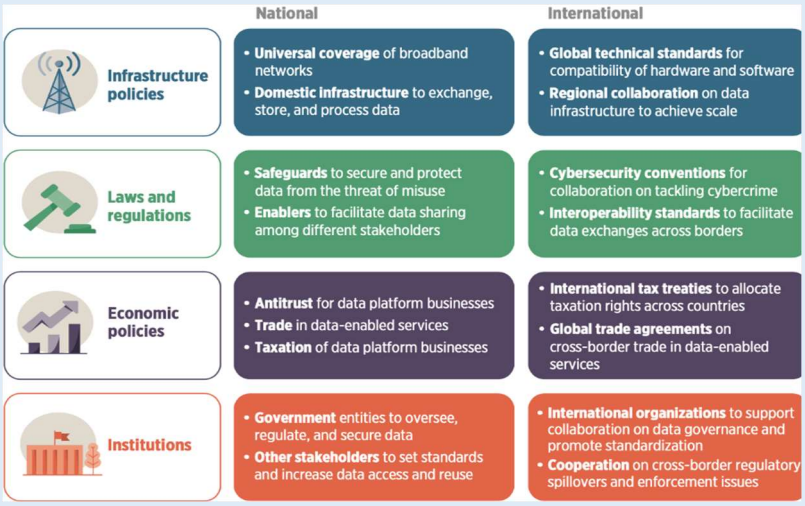
- **Trust** – a trust environment is created when the rights and interests that all stakeholders have in data are safeguarded.
- **Equity** – all share equitably in the benefits of data when investments/regulations create a level playing field.
- **Value** – the full value of data materializes when systems enable the use and reuse of data for different purposes.

The report identifies data governance as the most important instrument for establishing, facilitating, and enforcing social contracts for data. The three pillars together form the basis of an "Integrated National Data System", or INDS (Figure 6).

In conceptual terms, an Integrated National Data Ecosystem:

- Is an all-embracing data system involving people, processes, and technology for the sharing and reuse of a country's data resources at the national, urban, or sectoral levels.
- Aims to deliver digital dividends, ensure continuity of business and government services, and empower civil society.
- Provides easy, unified, transparent, and secure access to trustworthy, high-quality data including administrative, statistical, business, industrial, scientific, and real-time machine generated data.
- Consists of a set of interconnected, sector-specific trusted data platforms that are interoperable, where personal and non-personal data are secure; where stakeholders can freely exchange data subject to rules that guarantee data privacy; and where users have fast, reliable access to relevant information and services based on international and national data standards, and interoperability frameworks.

Figure 6. An Integrated National Data System



Source: World Bank WDR, 2021.

- implemented with state-of-the-art digital infrastructure including data stacks (preferably open source[47]), ethical software tools (data analytics, AI), enhanced cybersecurity and pooled distributed infrastructures (cloud, edge, IoT, networks) for managing and processing data to create value.

[47] Some potentially useful links to open-source platforms: World Bank Geospatial Platform: https://maps.worldbank.org/; KIDS (Knowledge, Information & Data Services) Helpdesk catalog: https://spatialagent.org/KIDS/; HydroInformatics eBook: https://spatialagent.org/HydroInformaticsEbook/; Open Data Platform: https://www.spatialagent.org/HydroInformatics/; GEOGLOWS-ECMWF Platform: https://geoglows.apps.aquaveo.com/apps/geoglows-hydroviewer/ and open services: https://geoglows.ecmwf.int/ for streamflow estimation; Open data from FAO: https://earthmap.org/

- Is underpinned by laws and regulations to establish trust and safeguard users' rights, facilitated by economic policies to support innovation and overseen by inclusive governance institutions, which are essential to ensure the quality, accessibility, protection, availability, reusability, and preservation of the country's unstructured and structured data and associated metadata.

While many countries have long had policies that governed the collection and use of data within their national borders, the introduction of the European Union's General Data Protection Regulation (GDPR)[48] marked a watershed moment for data policy around the world. Following the example of the GDPR, many countries have established laws, regulations, and institutions focused on individual privacy and the obligations of "data processors" to ensure adherence to principles such as data minimization, purpose limitation, data accuracy, storage limitation, data security, and accountability. India too has proposed a new Personal Data Protection Bill[49], which is currently under discussion.

***Data for Better Lives* provides a useful summary of many national policy regimes and can serve as a resource.** The report suggests there is a growing interest in broadening the scope of data policy within governments to include "ecosystem" issues like innovation and entrepreneurship, market conditions, capacity development, and engagement to create conditions for data-driven solutions to be most effective.

KERALA

The WB team has separately provided feedback and action items based on its review of the Kerala state data policy documents referenced earlier in the section. It is unclear why multiple policy documents have been developed simultaneously. The main themes of the feedback include the following suggestions:

- **Make the policies people-centered.** The current drafts are strongly focused on IT and/or technical elements of data policy; alongside the proposed work on a data strategy, the policy documents can serve as instruments to create data policy that is important from the perspective of citizens and businesses. Such a shift would increase the attention paid to "human" issues like capacity development, engagement, co-development, innovation and entrepreneurship, and data rights. It is also important to spotlight issues such as algorithmic bias that can affect the quality of decision-making in government and erode trust (data underlying algorithms is often a major cause of bias[50]).
- **Focus on data sharing and reuse**. The current policy documents, rightly, seek to clarify issues such as data classification and data processing obligations within government. It is however also important to emphasize value-creation through data. The separately shared suggestions include ideas to make data sharing more effective and seamless and to promote usage by a wide range of internal and external stakeholders (including machine readability and clear licensing).
- **Establish standards and protocols for data management within the state**. The WB team has made several suggestions on standardizing different aspects of data management in the state, including metadata practices, data classification, and data access provisions and processes. The WB team has also held preliminary discussions about global best practices for data standards that the team in Kerala can adopt.
- **Clarify institutional arrangements.** This has been reflected earlier in this report.

[48] Available at https://fra.europa.eu/sites/default/files/fra_uploads/fra-coe-edps-2018-handbook-data-protection_en.pdf.
[49] Available at https://www.meity.gov.in/writereaddata/files/The%20Digital%20Personal%20Data%20Potection%20Bill%2C%202022_0.pdf.
[50] Prasanna Lal Das. 2022. "Algorithms in government: A magic formula or a divisive force?" Available at https://dial.global/research/algorithms-in-government-a-magic-formula-or-a-divisive-force/.

4. Develop a plan to manage data staffing in the state government

CONTEXT

Data skills are in high demand and governments around the world, including in advanced economies, struggle to attract and retain digital talent. Kerala is no exception and most agencies consulted for this diagnostic reported challenges in sourcing IT skills. For entities such as KSITM that have been mandated to potentially serve as nodal entities for data innovation and data/digital transformation of government, or like KSDMA that must provide sophisticated digital services and tools as part of their core operational model, the skills challenges are even greater.

LEADING PRACTICE

There is no magic bullet to resolve the staffing challenge for digital skills in government – virtually no government can offer compensation or career opportunities for most staff that are on par with the private sector. The GoK may however explore one or more of the following staffing models that other governments have tried:

- **Bring external data talent into government (as staff) for high-value assignments**. Many countries have begun to experiment with staffing models that allow successful entrepreneurs and successful technologists to contribute to public sector initiatives from within the government. Examples that the GoK might consider include:
    - Digital fellowships with government. Data or digital fellows are data experts hired from the private sector or academia for short- to mid-term engagements (typically 12-24 months) to work on mission-critical problems while embedded within specific government agencies. The fellows contribute as advisors or hands-on technical specialists, and bring much needed external perspective and skills to government programs that otherwise cannot afford to hire such specialists. The Presidential Innovation Fellowship program[51], started by the US federal government in 2012, is a strong model for such programs. Other data fellowship programs include Baltimore's City Data Fellows[52] program and the Data Science Fellowship[53] program in many African countries. In Kenya, with the support of the World Bank, the Open Institute has been working with various subnational governments to develop the foundations for consistent use of data to strengthen participatory processes at local levels. Kenya's work on subnational data desks has set the stage for embedded data scientists to create efficient, problem-orientated service delivery.[54]
    - Joint development programs for services and solutions with the private sector and academia. Several co-development approaches have begun to emerge[55], which help governments develop digital public infrastructure (DPI)[56], strategic solutions, and services in collaboration with the private sector. The GoK may consider these as a solution to attracting personnel with the skill to develop and deliver digital transformation initiatives. The India Stack is a leading example of a government platform that provides the infrastructure for the private sector to develop applications that address market and service gaps.[57] The useful registry of Digital Public Goods (https://digitalpublicgoods.net/registry/)

[51] See https://presidentialinnovationfellows.gov/apply/track-ai-data-and-analytics/.
[52] See https://www.baltopi.com/data-fellows.
[53] See https://www.data4sdgs.org/initiatives/data-science-fellowship-program.
[54] See https://blogs.worldbank.org/opendata/subnational-data-desks-set-stage-more-efficient-participatory-development
[55] See https://www.codevelop.fund/enabling-governments-to-build-and-sustain-digital-public-infrastructure.
[56] See https://digitalpublicgoods.net/blog/unpacking-concepts-definitions-digital-public-infrastructure-building-blocks-and-their-relation-to-digital-public-goods/.
[57] See https://www.strategy-business.com/article/What-the-world-can-learn-from-the-India-Stack.

may be very relevant to the state's resilience initiatives. The proposed Unified Registry in Kerala has the potential to grow into a similar example of DPI.

- Partnerships with civil society organizations to serve as conduits for talent. Many governments have reported success working with civic-minded organizations that are deeply networked with digital professionals to identify and attract talent. The Government of Canada, for instance, worked closely with Code for Canada to secure placements for an incoming cohort of fellows.[58]
- Establish an on-demand pipeline of external talent for project hiring. Government processes and vetting requirements can make it difficult for government agencies to access external talent. GC Talent Cloud[59] in Canada is an innovative response to the problem, providing a curated list of vetted talent that any government agency can use for project-based work.

- **Expand career opportunities for data talent within government.** Many countries are creating opportunities for technologists, including data specialists, to have a more dynamic career within government. Options include:
  - Leadership development pathways within government. Apart from creating new positions such as Chief Data Officers, the Singapore government has established a Talent Leadership Committee[60] that systematically grooms data leadership talent within government and creates opportunities for them to serve in multiple agencies. The government has also created new data/digital leadership positions to ensure that there is representation of technical skill in all major strategic decisions.
  - Establish new types of talent marketplaces that enable greater mobility across government. It can be notoriously difficult for government staff to move across agencies and apply their work within multiple sectoral contexts. This is a particularly critical issue for digital experts whose work tends to have cross-cutting relevance. The Free Agent model[61] introduced by the Government of Canada is an attempt to make it easier for government staff with high-value skills to take control of their career paths. Not every public servant is eligible to be a free agent. Indeed, a key feature of the Free Agent model is the fact that it screens and selects public servants for their attributes and behaviors rather than their traditional educational credentials. The program is currently in a pilot stage and will be useful to test to determine whether these attributes will be valuable for project-based work and whether they will have an impact on problem solving and innovation in the Public Service.

- **Retrain public sector staff in data competencies and invest in continuous upskilling.** Government jobs have a reputation, sometimes deserved, of blunting technical skills over time, and many surveys suggest that the fear of eroding skills is one of the reasons many potential employees find government jobs unattractive.[62] Many governments have thus introduced internal learning programs – in the UK, for instance, the Government Digital Service (GDS) supports a range of programs such as the Tech Partnership[63] (network of employers focused on developing digital skills) and DotEveryOne207[64] (an independent think-tank focused on the digital society) as ways of broadening access to data skills development.

[58] See https://codefor.ca/uploads/pdf/Building-Digital-Government-Talent-Pipelines-Recruitment-and-Retention-for-Digital-Era-Government-Code-for-Canada.pdf.
[59] See https://www.oecd.org/gov/innovative-government/Canada-case-study-UAE-report-2018.pdf.
[60] See https://www.csc.gov.sg/articles/digital-government-smart-nation-pursuing-singapore%27s-tech-imperative.
[61] See https://wiki.gccollab.ca/Canada%27s_Free_Agents/FAQ.
[62] See https://www2.deloitte.com/us/en/insights/industry/public-sector/talent-war-government.html.
[63] See https://www.thetechpartnership.com/.
[64] See https://dotevervone.org.uk/.

KERALA

In addition to action steps along the lines of the ideas discussed above, the GoK may consider creating a center of excellence for data governance within the government that can provide a range of relevant skills. These skills would go well beyond IT/technology to help other government agencies better utilize and grow their data assets. The center may be housed within the institutions discussed previously. Examples from other countries include the UK's GDS[65] (see Box 2 for more on the UK experience), the US Digital Service[66], and the Government Technology Agency in Singapore.[67] Common services these agencies provide include technology platforms, service design guidance, procurement controls and assurances, and legal and regulatory advice. The skill mix for such entities often includes the following[68]:

- Strategists – staff who can provide a long-term vision and can craft and prioritize action plans to achieve related goals. In KSITM, the agency head must play this role.
- Domain experts – staff with deep expertise in relevant subject areas (e.g., resilience, water, transport).
- Technologists – staff with deep technical skills, both at the architecture and the coding levels.
- Design – staff with experience in service design, including user research and the user experience (this could include anthropologists and psychologists – i.e., experience technologists).
- Program managers – staff that can run a program, and manage people, resources, and reporting.
- Legal – staff for compliance and grievance redress.
- Engagement/communication – staff to focus on stakeholder participation, and creating awareness on services/tools.

Box 2. Snapshot: The United Kingdom's approach to building data skills in the public sector

In 2021, the UK government identified data as an area for reform[a], and launched a new curriculum to build skills in the public sector, including on data and data protection. Moreover, one of the objectives of the UK's National Data Strategy is to build data skills across the public sector, including by training analysts in data science.[b] Other examples of government attention to the skills issue include the Government Analysis Function, established in 2018, which hosts "Analysis in Government Month" to highlight relevant work and facilitate community building among government analysts.[c] "Analysis in Government Awards Celebration Day" was created in 2020 to recognize "excellence in analysis" within the public sector. The analysis function also has a learning curriculum that claims to offer something for everyone, even for those who work outside analytical roles. Furthermore, the government's "policy profession"[e] has a competency framework with a set of standards[f] that requires professionals to be able to "apply research methods to model, test and improve policy solutions" by using data and evidence. For additional details, refer to "Data Literacy and the UK Government."

[a] See https://www.gov.uk/government/publications/declaration-on-government-reform.
[b] Available at https://www.gov.uk/guidance/national-data-strategy.
[c] See https://www.gov.uk/government/organisations/government-analysis-function.
[d] See https://analysisfunction.civilservice.gov.uk/learning-development/learning-curriculum/.
[e] See https://www.gov.uk/government/organisations/civil-service-policy-profession.
[f] https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1065808/UPDATED_PP_Standards_main_v5_acc.pdf.

[65] See https://www.gov.uk/government/organisations/government-digital-service.
[66] See https://www.usds.gov/.
[67] See https://www.tech.gov.sg/
[68] As indicated as an area of interest by different representatives of GoK considering their staffing plans for the next three to five years. This is a partial list. Many data innovation teams also incorporate other roles such as IP experts, innovation advocates, etc.

5. Grow the demand side of the state data ecosystem

CONTEXT

Data has no value if it is not used. Government data initiatives around the world often fail to monitor the use of data resources, or discover that the data they have made publicly available is of little interest to citizens or businesses. It is also not uncommon for citizen engagement initiatives to fail, even when citizen participation is built into a program. For example, one study found little evidence of affected populations playing a significant role in the design or management of predictive analytics in the humanitarian sector.[69]

The GoK experience fits a common pattern. The state's open data portal[70] has been dormant since its launch and the datasets, which have not been updated since launch, are barely accessed or used. Interviews suggest that administrators of most other online platforms in the state do not collect detailed data about users and usage and have only anecdotal examples and evidence of demand for data by other government agencies. Almost no agency appears to track how its data is used and how it can engage with and support citizens, application developers, data scientists, and entrepreneurs/businesses.

There are exceptions – KSITM has run a series of successful mapathons[71] that have served as potent tools to engage local communities, develop digital skills, increase civic participation, and develop new mapping assets. Similarly, KSREC has developed a mapping application called Gramam that relies on local volunteers to supplement mapping data. KSDMA worked with a large number of data crunchers who volunteered their time during the pandemic to complement the physical infrastructure asset information in Google. Interviews also indicated that the state has organized several hackathons to engage students. The Kerala StartUp Mission works closely with entrepreneurs, a growing number of whom are developing digitally-enabled products and services.

LEADING PRACTICE

Key trends in evaluating and stimulating the demand for data and data-driven services around the world include:

- **Open data publishing standards and guidelines.** Instead of publishing data opportunistically (based on what is most expedient to publish), many jurisdictions start by identifying core datasets that are in demand by stakeholders and publish them in formats that encourage reuse. The EU recently published what it calls "high-value datasets" held by the public sector[72] in six data categories that are extremely relevant to the resilience agenda – geospatial, earth observation and environment, meteorological, statistics, companies, and mobility. The list also includes core data attributes and publishing format suggestions.
- **Metrics.** Governments use different metrics to establish the demand for and usage of data. For general purpose analytics, the Netherlands recently commissioned a study[73] of three open data portals where usage was tracked through the following tools (these serve as a useful proxy for techniques many governments use):

[69] Kevin Hernandez and Tony Roberts. 2020. "Predictive Analytics in Humanitarian Action: A preliminary mapping and analysis." Institute of Development Studies. Available at https://opendocs.ids.ac.uk/opendocs/bitstream/handle/20.500.12413/15455/EIR33_Humanitarian_Predictive_Analytics.pdf?sequence=1&isAllowed=y.
[70] Available at https://kerala.data.gov.in.
[71] See https://mapathonkeralam.in/.
[72] See https://digital-strategy.ec.europa.eu/en/news/commission-defines-high-value-datasets-be-made-available-re-use.
[73] Charlotte van Ooijen, Nathan da Silva Carvalho, Alice Iordache, and David Osimo. 2021. "Measuring data demand within the public sector." Data.europa.ed discussion paper. Available at https://data.europa.eu/sites/default/files/report/Discussion_Paper_Measuring_Data_Demand_Within_the_Public_Sector.pdf.

- Analyzing IP addresses and web statistics to identify user categories and countries, the data they use, data access methods, and the path taken toward the desired data sets.
- Web crawling techniques (e.g., search APIs) to identify websites that mention the reuse of open data.
- A quantitative analysis of Tweets mentioning open data.
- Online questionnaires and qualitative interviews with data re-users.
- Crowdsourcing the priorities, challenges and needs of data re-users

- **Data intermediaries.** Often defined as entities that manage relations, formally or informally, between data rights holders (individuals or agencies that release data) and data consumers (public or third-party applications), data intermediaries can be a powerful channel to help increase data use. One key reason to work with intermediaries is that many already have credibility within social and business communities. Other intermediaries can provide expertise that the government does not have. Examples of data intermediaries in action include:
  - Data journalists. Journalists are powerful interpreters of data; data-driven stories help people make sense of trends and relate them to their immediate environments. Newspapers such as the *New York Times*[74], the *Economist*[75], and the *Guardian*[76] are noted for their strong data-driven stories. A recent case study in the US on the impact of data journalism showed how data has been used to make government services such as property taxes fairer. Similarly, journalists in the Nile Basin countries have used data-driven journalism to prompt government action or policy intervention. In one case, authorities acted to protect the ecosystem of a six-acre island in Uganda that was being affected by floods and rising water.[77] Thus in many developing countries, where journalists in traditional news or media outlets are not always data savvy, governments have established programs to train journalists on leveraging data (see an example from the Africa Data Hub[78]; the WB has also offered data literacy programs for journalists in many countries[79]).
  - Data literacy. Increasing civic data literacy is a key way to encourage data usage and support data innovation in the public sector.[80] At the most basic level, Switzerland has experimented with a data literacy program in which young children teach digital fundamentals to older people. More ambitiously, in 2018, Finland announced that it will teach the basics of AI to 1 percent of its population – around 55,000 people who will study machine learning and neural networks.[81] The initiative was first started as a free online university course and over 250 companies, in addition to government organizations, have signed up to train their employees in AI.
  - Data rights groups. There is often a trust gap when it comes to using online services and many individuals/communities do not understand their data rights, or the obligations of governments and companies to respect their privacy and protect their data online. Activist

[74] See https://open.nytimes.com/tagged/data-journalism.
[75] See https://medium.economist.com/tagged/data-journalism.
[76] See https://www.theguardian.com/media/data-journalism.
[77] Kumar, Ravi and Ashlin Simpson. 2022. "Building Data Literacy and Facilitating Data Use in Africa: Highlights from the Partnership Between Code for Africa and the World Bank." Available at https://documents.worldbank.org/en/publication/documents-reports/documentdetail/099744310142241052/idu0727bdcf602040a86b055a85cb4fcff.
[78] See https://www.africadatahub.org/opportunities/odipodev-data-journalism-fellowship.
[79] See https://www.worldbank.org/en/programs/data-use-and-literacy-program.
[80] OECD. 2017. "Core Skills for Public Sector Innovation." Available at https://www.oecd.org/media/oecdorg/satellitesites/opsi/contents/files/OECD_OPSI-core_skills_for_public_sector_innovation-201704.pdf.
[81] See https://www.elementsofai.com.

groups, like the SHARE Foundation[82] in Serbia, for instance, can play an important role to help develop trust and to support the government on using data responsibly.

- Data trusts/collaboratives/fiduciaries. Many new data sharing models, designed to shift control of data from corporations and governments to individuals and communities, have begun to emerge to address the widely held belief that the data ecosystem is currently designed to promote the interests of data product developers rather than citizens. India is among several countries exploring new instruments and models such as data trusts/fiduciaries/collaboratives.

- **Engagement.** Governments around the world employ a range of engagement channels to reach and partner with citizens, businesses, and national/international agencies. Typical components of engagement strategies include data sharing, co-development and collaboration, representation and voice channels, and continuity (versus ad hoc or sporadic initiatives). Examples include:
  - Many governments run their own programs to reach citizens, employing tools such as hackathons and data dives, but intermediaries with well-established community roots, as described above, are often a more potent channel. One powerful example of the vital role that intermediaries can play in building trust in the data ecosystem comes from Moldova, where the Positive Initiative[83] uses several data tools to advocate for, and facilitate health services to, highly vulnerable and marginalized populations including AIDS patients, sex workers, drug users, prisoners, and refugees (marginalized groups who have traditionally been "unseen" by public services).
  - At the international level, Joinup[84] is a collaborative platform created by the European Commission that offers several services that aim to help e-government professionals share their experience with each other, with the goal of ensuring seamless and meaningful cross-border and cross-domain data exchanges between European public administrations.
  - Recognizing the vital role of businesses in data collection and use, governments around the world have invested heavily in digital innovation and entrepreneurship. The Digital Innovation Hubs in the EU[85] are a classic example of such an approach.

- **Co-development.** The idea of co-development stems from the recognition that:
  - All stakeholders should have input into the development of solutions that affect them. This is especially true for digital services, because the flow of data across the service lifecycle removes many of the distinctions between providers and beneficiaries.[86]
  - Governments can/should encourage collaboration instead of competition among its suppliers, in the right context. Recognizing the inefficiencies, knowledge gaps, and (risky) dependencies created by relying on individual suppliers for critical services, many governments (e.g., the UK[87] and US[88]) have begun to encourage co-developed solutions whereby organizations "voluntarily work together as equals to develop a mutually beneficial and jointly-owned product, service or other common objective".

[82] See https://sharefoundation.com/.
[83] See https://positivepeople.md/.
[84] See https://joinup.ec.europa.eu/.
[85] See https://digital-strategy.ec.europa.eu/en/activities/edihs.
[86] The theory here often outstrips practice. For instance, a recent study found very little evidence that community participation had been effective in reducing the risk of algorithmic bias in applications despite significant engagement effort: https://opendocs.ids.ac.uk/opendocs/bitstream/handle/20.500.12413/15455/EIR33_Humanitarian_Predictive_Analytics.pdf.
[87] Hiblin B, Calvert T, Hopkinson L, Van Ry R, Sloman L and Cairns S. 2021. "The Co-development Process: National Evaluation Case Study 1, Transforming Cities Fund Report to Department for Transport." Available at https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1006470/tcf-co-development-process-national-evaluation-case-study-1.pdf.
[88] See https://www.usaid.gov/co-creation-usaidm.

- **Digital public infrastructure (DPI) and digital public goods (DPG).** Increasingly governments recognize that the current infrastructure model of the Internet concentrates power in the hands of a small set of powerful technology and service providers. The DPI model envisions public digital rails that provide the foundation for others to develop products for large populations.
    - There is a growing need to create new shared infrastructure and public services that can be used by a wide range of stakeholders. The IndiaStack[89] is a classic example of foundational infrastructure that provides services to private and civil society actors who, in turn, build people-facing solutions. The new data spaces in the EU, as described later in this report, are another powerful example.
    - A recently published report from the World Bank provides a useful list of DPI resources for resilience.[90]

KERALA

Consistent with the examples above, the GoK should consider the following suggestions to identify and increase the demand for data in the state and engage systematically and deeply with a wide range of stakeholders in the local data ecosystem (see also Box 3 for experiences from Korea):

- **Publish data strategically, and for reuse.** The GoK currently lacks a well-defined or targeted publication process, and much of the data it publishes is in a format that makes it difficult to reuse the data (currently some of the spatial data is published by the state in "dumb" formats, for instance, that do not encourage reuse). The GoK should initiate an inventory and gap analysis of its data holdings benchmarked against the EU list of "high value datasets" referenced above, which would create a clearer picture of the quality and coverage of the data it publishes.
- **Better measure the impact of data initiatives.** There is currently no systematic analysis of data use on government websites beyond high level numbers indicating traffic and downloads (which are not always tracked). This means the government does not know the impact (or lack thereof) of its data initiatives. Impact measurement can be notoriously difficult, but useful resources are available.[91]
- **Identify and work with data intermediaries with strong local credibility and grassroots presence**. A strong decentralized system is one of the core strengths of the governance model in Kerala, but the government is not geared to work closely with communities on local data issues and opportunities. This is where data intermediaries come in. As described above, there are a variety of data intermediaries – ranging from data journalists and data rights advocates to formal entities such as data trusts and data fiduciaries. The GoK should identify and support data intermediaries that can help advance its development agenda with local populations.
- **Widen and systematize the range of engagement channels to reach different types of audiences.** It is true that the government already employs several engagement channels to promote its data products and services, including consultations and workshops and formal events such hackathons and mapathons. Feedback channels also exist on several websites. However, these engagement channels are not used systematically or in coordination with each other and there is little targeted measurement of such activities. The online feedback channels do not always appear user-friendly (they can be difficult to locate on many websites, for instance). The diagnostic team could not ascertain the effectiveness of social media use.

[89] See https://indiastack.org/.

[90] Global Facility for Disaster Reduction; United Nations Office for Disaster Risk Reduction. 2023. Digital Public Goods for Disaster Risk Reduction in a Changing Climate. http://hdl.handle.net/10986/39551

[91] This website provides a strong starting point: https://datos.gob.es/en/blog/measuring-impact-open-data.

- **Explore co-development models to create products and services.** The government should explore opportunities to jointly develop data-driven products and services with non-governmental entities who can participate either as developers or contributors to the design process. Co-development helps embed demand/user perspective in solutions that can otherwise tend to be supply-oriented.

Box 3. Snapshot: The Republic of Korea's experience responding to data demand

A Korean high school student in 2009 used municipal data to create an app called "Seoul Bus" to notify users of bus routes and arrivals in real time. However, the app was suspended on the grounds that it was an unauthorized use of government data.[a] Citizens protested the government's move to restrict this use of data, which had been intended to improve a public service, and the app was made publicly available again. This incident made use and reuse of open government data a matter of policy discussion in South Korea. In 2013, the government enacted the "Act on promotion of the provision and use of public data".[b] Over the last decade, South Korea has taken several steps to respond to public demands for data, and to actively foster its use to improve public services, strengthen resilience, and enable the creation of new businesses. The government has created "Open Square-D" in cities such as Seoul, Busan, and Gangwon to provide office space for entrepreneurs to exchange ideas and develop products. For example, an app called "GoodDoc" lets users search for nearby hospitals and make reservations. The app has been downloaded over 5 million times. "Kweather", an app that provides weather forecasts and fine-dust forecasts using real-time API data, enables citizens to track weather conveniently and make necessary preparations in advance of rain or storms. Korea's experience shows that when public data is released, used and reused by diverse stakeholders, it has the potential to contribute to economic growth and improve lives.

[a] See http://www.koreaittimes.com/news/articleView.html?idxno=58369
[b] See https://elaw.klri.re.kr/eng_mobile/viewer.do?hseq=47133&type=part&key=4#:~:text=The%20purpose%20of%20this%20Act%20is%20to%20prescribe,of%20such%20public%20data%20in%20the%20private%20sector
[c] https://www.data.go.kr/en/osd/opensquare.do
[d] https://play.google.com/store/apps/details?id=com.ksncho.hospitalinfo&hl=en_GB&pli=1

Additional strategic themes

This section describes two additional strategic themes that the GoK should consider to broaden the scope of its data initiatives. The suggestion is to approach these in an exploratory way, thinking of them as frontiers of the data agenda in comparison to the concrete actions proposed above. They reflect emerging trends and approaches that can potentially be critical for the success of the state's data initiatives in the longer term.

1. Develop a "data space" for resilience in the state

Resilience is the classic boundaryless modern sector. It draws on expertise, infrastructure, and resources from multiple traditional sectors (such as water, energy, transport, and agriculture); it is deeply intertwined with other multidisciplinary, global challenges like climate change; it operates across political and physical boundaries; and it requires a whole-of-society approach that obliges governments (at all levels), businesses, civil society organizations, academia, and individuals on the ground to cooperate, collaborate, and share resources to achieve common goals. The data strategy to strengthen resilience must therefore be unencumbered by organizational boundaries and flexible enough to allow multiple actors to leverage and contribute to shared resources that can be deployed locally or at scale in a wide variety of contexts.

The concept of "data spaces" as described in the 2019 European data strategy[92] may be a useful starting point for the state of Kerala to consider as it develops its own statewide data strategy and simultaneously identifies/establishes policies and infrastructure for resilience. The document defines a data space as "data infrastructure with tailored *governance mechanisms* that will enable secure and cross-border access to key datasets in the *targeted thematic area*." Key features of a data space[93] include:

- A secure and privacy-preserving infrastructure to pool, access, share, process and use data.
- A clear and practical structure for access to and use of data in a fair, transparent, proportionate and/non-discriminatory manner, and clear and trustworthy data governance mechanisms.
- European rules and values (or in the case of Kerala – local rules and values) – in particular, personal data protection, consumer protection legislation and competition law are fully respected.
- Data holders will have the possibility, in the data space, to grant access to or to share certain personal or non-personal data under their control.
- Data that is made available can be reused with compensation, including remuneration, or for free.
- Participation of an open number of organizations/ individuals.

The EU Data Governance Act[94], approved in May 2022, is expected to "pave the way for sectoral European data spaces to benefit society, citizens and companies". The European data strategy originally envisioned nine such data spaces (the number has grown since then) ranging from manufacturing to energy, and from health to the European Green Deal (Box 4). The WB can share additional details about the implementation status of data spaces and arrange knowledge sharing events if the GoK is interested in learning more about the European experience. It is also useful to note that the use of new forms of data and data-related analytics and infrastructure in traditional sectors is emerging as a clear theme in the digital strategy of several countries[95] both in the developed and developing worlds. These may also be themes for future knowledge exchange.

Box 4. Snapshot: The EU Green Deal Data Space

The Green Deal data space in the EU focuses on resilience and sustainability. The data space serves as a platform for government and non-government actors to share and jointly develop project ideas and utilize a shared infrastructure to build tools that can be deployed to meet the EU's Green Deal objectives. The platform currently provides the following modules to solution developers:

- A data marketplace containing 2 million open datasets.
- An open innovation space to organize "datathons" focused on specific challenges.
- An AI Workbench that provides a range of data processing tools.
- Secure data sharing in "closed user groups" designed to facilitate the use of sensitive data that cannot be shared on the data marketplace.
- A trusted data hub designed to address data providers' concerns about losing control of their data (the hub provides tools so "even sensitive data can be used collaboratively by multiple parties with full data sovereignty and preservation of the confidentiality of the raw data").
- A connection module to encourage community building.

The service is currently in early days, and it is difficult to assess effectiveness. Questions remain about design, technology, and community engagement, but it provides a good model for Kerala.

[92] Available at https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52020DC0066.
[93] See https://digital-strategy.ec.europa.eu/en/library/staff-working-document-data-spaces.
[94] See https://www.consilium.europa.eu/en/press/press-releases/2022/05/16/le-conseil-approuve-l-acte-sur-la-gouvernance-des-donnees/.
[95] Prasanna Lal Das. 2022. "In machines we trust? Data governance and the transformation of traditional sectors." Available at https://dial.global/research/spolight-paper-machines-trust/.

2. Incorporate more use of machine data in service design and delivery

Machine-generated data has already outstripped human-generated data in size, variety, and frequency and is growing at a staggering pace. According to one estimate, there is 50 percent compound growth in the size of data flows every year, and there will be 125 billion networked devices by 2030.[96] The growing prevalence of Internet of Things (IoT) devices is leading to the increased "datafication" of traditional sectors[97] such as agriculture, transport, energy, natural resources, and health – each of which is a key contributor to the resilience agenda. The GoK has already begun to deploy a variety of sensors in its infrastructure – examples include cameras, but also sensors within public infrastructure such as roads and water pipes – but there appears to be limited capacity within government to either deploy, maintain, or use these sensors strategically to support effective decision-making. Interviews also suggest, though the diagnostic team has been unable to verify it, that many sensors are not appropriate for the task or do not work as intended.

Given that machine data is going to be one of the dominant sources of data (and frequently of algorithmic responses) for the resilience agenda, the diagnostic team recommends that the proposed state data strategy should include machine data as a core component. Meanwhile, there are a few tactical steps that the GoK should consider to explore the usefulness of machine data at scale to deliver public services (based on practices observed in other jurisdictions, such as Estonia, Germany, and Kazakhstan[98]):

- **Establish a sandbox program to test the use of machines to develop/deliver resilience applications and services.** Sandboxes are physical spaces, clusters, and/or environments for running pilots and proofs of concept. Sometimes referred to as living labs or model cities, these physical spaces – such as buildings, streets, or bridges – provide facilities for setting up test beds and building "models" ranging from simple IoT applications to even "model" cities. The proposed Urban Observatory in Kerala can incorporate many features of sandboxes observed in other countries.
- **Establish linkages with other governments/agencies with experience incorporating machine data in operational processes.** Several agencies in Kerala are already experimenting with machine data, as are other government and non-government actors in India and abroad. The government should set up mechanisms to foster learning among practitioners and policymakers that work on machine data.
- **Develop capacity-building programs within government to grow awareness of IoT among government officials (see also Box 5).** Managing IoT-based applications and processes requires a very specific skill set and competency. For example, remote regulatory inspections may not require physical observations but may call for strong analytical skills and capabilities. Several governments are already starting to engage and partner with universities and academic institutions to develop appropriate curricula, starting with early education and continuing through college/university studies. The GoK should consider a formal plan to introduce machine data-related competencies in the government. The government should also ensure that regulations are in place for the safe collection and use of machine data (including periodic data audits).

[96] See https://en.ctimes.com.tw/DispNews.asp?O=HK1APAPZ546SAA00N9

[97] Ibid.

[98] Prasanna Lal Das, Srikanth Mangalam, Mehmet Rasit Yuce, Stefan Claus Beisswenger, and Martin Lukac. 2017. "Internet of things: the new government to business platform - a review of opportunities, practices, and challenges (English)." Washington, D.C. : World Bank Group. Available at https://documents.worldbank.org/en/publication/documents-reports/documentdetail/610081509689089303/internet-of-things-the-new-government-to-business-platform-a-review-of-opportunities-practices-and-challenges.

Box 5. Snapshot: Priority passage for emergency vehicles, using IoT

India Urban Data Exchange (IUDX), along with the Vadodara Smart City, has developed a "Green Corridor" for ambulances. Data from Adaptive Traffic Signals (data that resides in the Internet Traffic Monitoring System – ITMS – the system in the command center for traffic lights) is collected using APIs, and the Health Department's GPS data from 100 ambulances (data from onboard GPS systems) is collated in real time. Whenever an ambulance comes within 300 meters of a traffic signal, the signal automatically changes to green, allowing the ambulance to proceed.

Tactical action steps

To support the strategic actions that have been outlined, based on the diagnostic the GoK should consider the following tactical steps (Table 2). These steps are a mixture of action items that are already underway – but in an uncoordinated fashion – and actions recommended to address gaps observed during the diagnostic. Please note that this is not a comprehensive list, and that the list is not in order of priority.

Table 2. Recommended tactical action steps

#	Action item
Steps to establish and formalize data governance practices	
1.	Develop an inventory of data collection tools and expand/retire tools based on gaps/duplication.
2.	Develop an inventory of resilience-related data assets and do a gap analysis against global practice (a useful starting point is the list of high-value datasets developed by the EU).[99]
3.	Formally track demand for GoK data, both from within the government and from outside – establish related metrics and identify analytical tools that can/must be used for measurement.
4.	Create an inventory of known data gaps within existing applications and develop a process to plug the gaps if relevant data already exists within other government agencies (add LSGD data to Kerala Water Resource Information System data, for instance).
5.	Similarly, formally inventory data held across departments and develop processes to ensure that existing data is reused (rather than re-created); WRD and KSREC, for instance, appear to have overlapping data collection activities.
6.	Establish reporting requirements for data sharing and use.
7.	Initiate the process of identifying and documenting the ownership of the "golden" source of each data variable that the GoK collects and uses across agencies. Conduct periodic data audits to ensure that the 'right' data is being used for decision-making.
8.	Provide interim guidance about spatial data that can be shared publicly and establish that the data must be shared in vector formats (ideally through an API – at least for low volume users).
9.	Start an awareness campaign among state departments about the state's data initiatives and data assets (so agencies do not duplicate efforts).
10.	Further strengthen the virtual cadre of data/digital specialists within the GoK that can serve as a community of practice and a data resource center for other government staff.
11.	Develop an engagement strategy with the private sector to share data, and co-develop solutions and applications; build capacity among local entrepreneurs and provide resources to them.[100]

[99] See https://digital-strategy.ec.europa.eu/en/news/commission-defines-high-value-datasets-be-made-available-re-use.
[100] The IT Mission provides a useful model for such an exercise.

#	Action item
12.	Establish state level Digital Awards modeled on the Digital India Awards to recognize innovation and reward teams for implementation.
13.	Learn to tell stories through data.
14.	Provide provisional guidance for and standardize data collection and cleaning.
15.	Initiate user experience research as the new citizen portal is developed and other government applications are refreshed.
16.	Develop a data and user experience style guide based on the user experience research and ensure that all new applications leverage these resources.
17.	Ensure that all agencies that supply GIS data to KSITM have identified/created proper GIS teams to collect/clean/share the data (currently Kerala State Spatial Data Infrastructure [KSDI] needs to constantly re-engage with new personnel from teams that supply GIS data).
Steps to coordinate/align ongoing data activities	
18.	Address known data sharing gaps. Examples include: • Share information on the fair value of land (set by the planning department) with the land revenue department to eliminate discrepancies in fair value assessment. • Share river water level data, collected by WRD, with the transport department in real time so it can take preemptive measures, if necessary, to avoid public inconvenience. • Share data related to building permits with the LSGD so it can issue permits more efficiently. • Explore opportunities for improved data sharing between the registration department and the town planning department. • Explore better data sharing related to occupancy software and taxation systems. A person is liable for tax payments as soon as an occupancy certificate is issued, but the taxation department is not notified of the issuance of occupancy certificates, which delays tax collection.
19.	Develop coordinates for all addresses – convert all government data into GIS format.
20.	Begin to develop an API infrastructure to share spatial and other data.
21.	Update the open data sets that are already available through the state open data portal, which has been dormant for the three years since its launch.
22.	Determine a licensing model to accompany state datasets and provide licensing information to all users of these datasets.
23.	Standardize boundary data that is used in all state applications and clarify the official source for the data (election commission?) – KSDI has recently mapped panchayat level boundaries and is awaiting government approval to publish the data. This is data KSDMA requires for multiple purposes.
24	Continue the process of digitizing all data managed by government. Many projects are already underway, and the government should aim for complete coverage.
Application-specific action items	
25.	Consider opening non-confidential ReLIS (Revenue Land Information System) data to the general public.

Use cases

As described previously, a key challenge for the GoK is fully leveraging its existing data. Moreover, when pilot use cases are developed, they are not always scaled or deployed effectively across agencies. KSDI, for instance, has developed an alert system for fire and other emergencies, but it appears that the Smart City program is developing another alert system instead of building on the KSDI pilot (the diagnostic team could not verify the full story).

The purpose of this section is to provide examples of data-driven tools in action, in jurisdictions around the world, that are intended to increase the resilience of companies, societies, and governments. Typical elements of a data-driven intervention include:

- **Application of "big data".** Typically, vast datasets derived from multiple public and private sources are used. The data is often collected in close to real time, from various sources including machines and online applications, in a variety of formats, in an unstructured way, and sometimes without the consent of the data source.
- **AI/machine learning/algorithmic decision-making.** Common tools are used to make sense of big data (typically by finding patterns using machine-learning techniques, without rules-based instruction from humans or under their constant supervision). Many algorithms replace or augment human decision making (with attendant opportunities and risks[101]).
- **Statistical modeling.** In a big data scenario, where the cost of developing individual models is very low, typically a combination of thousands of models is used to develop scenarios that anticipate events.

Examples in Table 3 are organized per the phases of the disaster management lifecycle – preparation, response, recovery, and mitigation (please note that this is a rough framework, and the reader is advised to consult other WB publications for formal and comprehensive definitions of core concepts).

Table 3. Data-driven tools in action

PREPARATION	RESPONSE
The preparedness stage is designed to improve readiness before a disaster occurs. Anticipation, hardening of defenses, and prepositioning of assets are core components of preparedness. 1. Data for public awareness of hazard risk, USA a. Provides users with local community-level awareness of hazard risk, exposure and vulnerability b. Users can visualize a community's combined physical exposure, socioeconomic vulnerability and markers of resilience to natural hazards c. At the census tract level	Response occurs when a disaster has hit. Core elements of response include warning systems, tracking of events and people, monitoring the state of the infrastructure, and effective and timely communication with people and responders. 1. Sentiment analysis plus user-generated data, Haiti a. To improve distribution of supplies among hospitals, rescue centers, and demand points b. Effective but also raised quality concerns 2. IoT technology and mobile cloud computing

[101] Prasanna Lal Das. 2022. "Algorithms in government: A magic formula or a divisive force?" Available at https://dial.global/research/algorithms-in-government-a-magic-formula-or-a-divisive-force/.

2. Sensors in railcars, Canada a. Designed not only to track trains, but monitor for changes in ambient conditions like temperature, movement, and pressure – and thus mitigate risks from hazardous, volatile cargo b. Communications protocols to alert responders	platforms, Taiwan and elsewhere a. For rapid evacuation of large populations b. Designed to achieve improved convergence rates, shortened evacuation route lengths and evacuation time, and balanced capacity in surrounding shelters
RECOVERY	MITIGATION
The recovery phase focuses on the stabilization of the situation on the ground. Typical activities include provision of benefits, humanitarian relief, repair/restoration of infrastructure, and resumption of "normal" activity. Surprisingly, this is less discussed in academic literature relative to other stages. 1. GPS and social data to map the movement of people, New Zealand 2. Blockchain, smart contracts, and Building Information Modeling (BIM) creates integrated, self-regulating, self-monitoring, and cyber-resilient transactional operations to automate the building permitting process in the recovery and rebuilding phase of a natural disaster	The mitigation phase focuses on reducing the chances of the recurrence of disaster (when possible), and lessening its impact on people, communities, and infrastructure. 1. Preventive maintenance of critical structures, Japan a. Sensors to regularly assess the integrity of critical structures b. Escalation to maintenance/repair crews when anomalies are detected 2. Detecting forest fires in Izmir, Türkiye a. Network of cameras monitors forested areas b. Machine learning used to detect forest fires when they occur c. Community-based warning system d. Integration with fire response

KEY FEATURES OF THE USE CASES

While each of the use cases has features specific to its local context, they do reveal a few important common characteristics relevant to the GoK:

- **Data flows across boundaries.** In all examples, data flows across organizational boundaries, with multiple agencies involved at different stages. In the example of forest fire monitoring in Türkiye, for instance, the cameras are installed and maintained by the city's IT department, which also manages the data, but the fire department has data access, as do communities (as required).
- **Tools are aimed at the ultimate user, the citizen.** There is a growing emphasis on creating tools that average citizens can use to make sense of data and make appropriate decisions. The hazard risk database in the US, for instance, includes powerful visualization and filtering tools for citizens to gather and track highly granular local data across a wide variety of parameters.
- **Experimentation and evolution.** Data-driven tools, despite their seeming ubiquity, are still new and their creators are still experimenting with their design and validating their effectiveness. Widely deployed sentiment analysis tools are a useful example of an approach that has thrown up unanticipated challenges related to disinformation, prioritization, and iniquity.

- **A premium on the user experience.** Disaster situations are high-pressure by definition and it is critical to ensure that the tools the government deploys are intuitive by default to all potential users, regardless of their data literacy levels and technical competence. In most of the examples above, the data initiatives were accompanied by intense community engagement during design and rollout, and testing of tools among affected populations, including relief providers.
- **A combination of offline and online components.** Offline tools and processes continue to be vital for online applications to be adopted and succeed. The examples above involve cooperation with offline participants (as segments of the population often are during emergencies).

Kerala Workshop

The WB mission team conducted a workshop in Thiruvananthapuram in January 2023 to discuss and brainstorm about ideas for resilience-related use cases with a group of government officials drawn from multiple agencies. The workshop generated several suggestions including the following (please note that these are not use case recommendations from the WB; they reflect the proceedings of the workshop):

1. New risk models for resilience in coastal towns:
    - Enable authorities to assess risks, develop risk models regarding the type/magnitude of risks, estimate probability that a disaster might occur, and estimate financial value of properties using geological and geospatial data, weather data, hydrological data, data on structure and infrastructure, population, property valuation, structural data, and town planning data.
    - Based on the model, communicate risks to the local community and develop a plan for compensation proactively.
    - Encourage settlements with highest risk to take compensation to relocate to safe zones.
    - Develop a multichannel communication system that includes text messages combined with engagement with community groups and feedback mechanisms via local officials to get feedback to KSDMA about the uptake of and response to forecasting.
    - Socialize guidelines for evacuations and organize drills proactively.
    - Develop guidelines and plan for evacuation and rescue.
2. Vulnerability assessment through citizen data collection:
    - Collect citizen-level data, by citizens & disseminated for citizens (participatory data collection owned by community) with bias correction. Overlay risk information by geographic location.
    - Inform citizens in real time about their vulnerabilities and options, which would allow them to make short-term and long-term decisions, including residency. This system should also have a feasibility analysis tool for citizens.
    - Data collected should be consolidated in a coordinated manner at the local level, to be available for local/state organizations.
3. A data demand prediction tool for the government:
    - A data tool for the government that predicts the datasets that may be needed for a particular task and suggests the departments that should be contacted.
4. A government demand-pooling application:
    - A system to analyze stakeholder interests (e.g., the electricity authority wants to store water in dams for energy production, while the water authority wants to release stored water to mitigate downstream flooding) and suggest potential outcomes/scenarios/compromises.

The use cases that emerged would need to be sketched out further before they could be implemented. Interestingly, most of the use cases that were discussed focused on challenges that government agencies face. There was relatively less reflection on the citizen demand side.

INTERVIEWS

Many use case ideas surfaced during the data diagnostic team's meetings with various stakeholders. Examples include the following:

- A life events and benefits tracking application:
  - Automatically stop pensions after the death of the pension owner. This would be a potential demonstration of the usefulness of the proposed Unified Registry, the analytical tools developed by SDMA, and the data in the KSMART registry.
  - Distribute benefits (sometimes proactively) in event of a natural disaster.
- A disease surveillance tool:
  - Identify vulnerable hotspots based on a variety of otherwise disparate data including weather forecasts, zoological observations, water quality, sanitation, and spatial data. The models currently in use are based on Excel sheets.
  - Integrate disease outbreak data with flood or heatwave related data to better understand the climatic links and mitigate effects.
- An infrastructure risk monitoring tool (or a set of connected tools):
  - Create vulnerability-linked relocation plans.
  - Provide inputs into the building permitting process.
  - Track road conditions, including accidents.
  - Integrate climate risk data into local body Disaster Risk Management plans.

POTENTIAL NEXT STEPS

Use cases do not emerge and cannot be successfully implemented and (when required) scaled without an adequate data governance framework to support them. Otherwise, most use cases are doomed to stay within innovation labs as experiments through which stakeholders learn, without achieving impact at scale. The GoK should consider potential new use cases in association with the strategic and tactical steps suggested above. A few techniques to consider for developing use cases include:

- **Design from the outside-in and engage stakeholders, especially the external ones, at the outset.** The government data ecosystem is still very insular and the government must meaningfully engage external stakeholders before initiating any new use cases.
- **Think beyond government data.** The GoK is actively engaged in collecting local data, but so are the private and civil society sectors (the private sector accounts for the bulk of data in most parts of the world). The government must develop data partnerships and shared tools with external data stakeholders as it develops new use cases.
- **Establish and leverage data innovation hubs that bring a range of data and resilience stakeholders together.** Innovation is a continuous process and the government needs to instill a culture of ideation and experimentation for viable ideas to emerge, be quickly tested, merged, and developed.
- **Engage the appropriate institutions.** The government needs to ensure it identifies and engages the right partners as it develops use cases. These may often be agencies that contain new areas of expertise or that create access to newer types of data.
- **Focus on people, not technologies.** Many government agencies fall into the trap of equating innovation with new technologies. The new government use cases must solve practical problems that communities face, with or without new technologies.

Appendices

Appendix 1. Main findings

The main findings of the rapid diagnostic are organized by the eight pillars of the ODRA methodology. Please note that this is not meant to be a comprehensive assessment and the findings below must be considered provisional.

Senior leadership

Highly visible leadership, backed up by an explicit policy/institutional mandate, is a significant critical factor for the success of data initiatives. The leadership environment in Kerala is mixed. There is a strategic mandate for data-led development, the state has some strong individual leaders with respect to the data agenda, and it has developed several well-received digital applications. However, the institutional leadership structure for data is still not fully developed and there is no functional leadership for the data agenda specifically right now. The main observations, based on consultations on the ground, include the following:

- There is support for and interest in the state's digital agenda from the top. Since May 2021, the Chief Minister's Office has been directly overseeing the state's IT portfolio. E-governance and digitalization were explicitly identified as priority areas for the state in the new government's very first cabinet meeting. The Chief Secretary of the state organizes weekly meetings with Departmental Chief Secretaries on e-governance initiatives. For example, officials in charge of digitizing government transactions, and implementing the Kerala Fiber Network[102] and the Unified Registry, provide weekly updates to the Chief Secretary of the state.
- The state is emerging as one of the country's public sector digitalization leaders. In June 2022, Kerala ranked first in the country in providing IT-based public services, according to the National e-Governance Service Delivery Assessment.[103] Several digital applications created in Kerala have received national recognition. Examples include the Ksheerasree Portal[104], the Digital Workforce Management System[105], and the Kottayam District Website[106], which have received Digital India Awards[107] for digital initiatives at the grassroots level, for collaborating with startups, and for complying with Guidelines for Indian Government Websites (GIGW) and accessibility guidelines, respectively.
- Interviews suggest that the GoK has been fortunate to have had strong leaders with a clear digital vision in charge of different agencies. These leaders have been able to help conceptualize and initiate many sophisticated digital programs; the leadership has however been individual rather than institutional and interviews suggest that most leaders leave a vacuum (and often do not transfer knowledge) when they leave.
- The institutional arrangement for leadership of the data agenda is unclear. KSITM has the official mandate for IT in the GoK and is deeply engaged in several core data initiatives in the state, including the proposed Unified Registry, management of the state data center and the spatial data infrastructure, and development of state data policy documents. It however does not have

[102] See https://kfon.kerala.gov.in/.
[103] See https://www.onmanorama.com/news/kerala/2022/06/14/kerala-ranks-first-in-e-governance.html.
[104] See https://ksheerasree.kerala.gov.in/.
[105] See https://knowledgemission.kerala.gov.in/.
[106] See https://kottayam.nic.in/.
[107] See h.ttps://digitalindiaawards.india.gov.in/.

unambiguous authorization to be the nodal data agency for the state (as described elsewhere in the document, the lines between data and other functional areas such as IT or e-governance are blurry). Consequently, other state agencies such as the KSDMA and LSGD have established their own data labs and teams, while others such as KSREC, IT Mission (LSGD), PWD, Centre for Development of Imaging Technology (C-DIT), National Informatics Centre (NIC), and others have developed many data applications apparently without any formal coordination with KSITM.

- The state has not appointed a formal CDO yet, but the position has been discussed in different ways in different policy documents. As envisaged in the proposed state data policy document, the CDO will serve as a data processor responsible for vetting requests for data sharing. The KDSAP would assign a wider set of responsibilities to the CDO. It envisions positions at the department/organization/state level (at the Head of Department level) and with wider responsibilities, including implementing the open data mandate and monitoring feedback from citizens and government functionaries. The two potential role descriptions should at a minimum be reconciled. It is also important to note that, as described elsewhere in the report, the CDO role in many jurisdictions tends to focus more on value from data than data management (as is the current plan in Kerala).

Policy/legal framework[108]

Establishing the ambit of the data policy framework in the state is tricky, especially when it comes to topics such as data privacy and protection for which national policies are still being finalized. While many foundational elements of data policy rest with the national government, the GoK has begun to formalize state level data policy through two main documents that are currently under consultation. These are the Kerala State Data Policy (KSDP-2022), developed by the Department of Electronics and Information Technology, and the Kerala Data Sharing and Accessing Policy (KDSAP - version 3.0), developed by KSITM. The two documents complement the Draft Kerala Spatial Data Sharing Policy (KSDSAP 2020), developed by KSREC.

Each of these documents draws from their equivalents released by the national government and reflects ongoing attempts by the state administration to bring order and clarity to its data policy environment, which is frequently marked by ad hoc and informal arrangements that rely on personal relationships and initiatives rather than systematic tools. The content of the policy documents also suggests a growing focus on data sharing arrangements within government, standardization of data collection, and development of institutional arrangements for state data governance.

The WB has shared comments on the two drafts separately. Related issues that emerged repeatedly in the consultations included:

- There is no clear policy establishing that data is a public resource that can be used and reused, and neither government agencies nor citizens are able to fully leverage data to make decisions, allocate resources, build applications to monitor weather more effectively, or inform communities about potential disasters that might be on the way so they can prepare.

[108] The preliminary analysis and recommendations in this section are based on information and opinions collected from interviews undertaken and materials provided by the government and other local stakeholders during this study. This section is not based on detailed, legal due diligence and does not constitute legal advice. Accordingly, no inference should be drawn as to the completeness, adequacy, accuracy or suitability of the underlying Diagnostic, or recommendations, or any actions that might be undertaken resulting therefrom, regarding the enabling policy, legal or regulatory framework for Data in the Kerala. It is therefore recommended that, prior to undertaking any legal action to address any legal Diagnostic issue raised herein, a formal legal due diligence be performed by competent, locally qualified legal counsel, preferably assisted by international legal experts with relevant experience and knowledge of these areas.

- In the absence of well-defined data sharing policies, the data sharing environment continues to be characterized by unclear and ad hoc processes. Most data sharing arrangements are informal and driven more by the personalities involved than by a formal process or set of requirements.
- Most interviewees expressed a commitment to open data but also pointed out many hurdles due to a lack of the following:
    - Clarity about which datasets should be open and the process to arrive at a decision.
    - Clarity around data formats and reusability.
    - Data updates or maintenance schedule.
    - Requirement to collect or use any usage analytics.
    - Metrics to measure success.
    - Formal definition or use of engagement channels.
    - Coordination with related open data initiatives.
    - Technical standards.
    - Clarity about licensing requirements.
    - Clear process on who should request data from whom.

Institutional structures/capabilities

As with data policy, the institutional structure supporting the state's data agenda overlaps with corresponding institutions at the national level. The following is a rough overview of the current institutional structure and the capabilities of various institutions and agencies in the state that either have a mandate for data-related activities or participate as key users/consumers of data.

There are several agencies in the GoK that have separate, sometimes overlapping mandates within the government data ecosystem. Here are some of the agencies that the diagnostic team consulted:

- **KSITM.** The Kerala State Information Technology Mission is the nodal IT agency for the state. It has historically been responsible for IT procurement and managed the GoK's key IT infrastructure, including e-office applications, all back-end support, and cybersecurity. It is currently drafting data policy documents for the state (discussed elsewhere), developing the Unified Registry application, and crafting the Digital Kerala architecture. It also manages KSDI, the flagship state geoportal. It is in some ways the nodal agency for the state's digital agenda, but its role is not universally understood as such by several other state agencies.
- **KSDMA.** The Kerala State Disaster Management Authority is the state's nodal disaster management authority. Data-driven products and applications are a core part of almost all its services, and it has established a data lab with the infrastructure and competencies necessary to deliver select data-driven services. KSDMA has developed several algorithms and models that may be relevant to other parts of the state machinery, but no formal platforms or channels are currently available for these services to be deployed at scale outside KSDMA.
- **KSREC.** The Kerala State Remote Sensing and Environment Centre is responsible for developing, updating, and maintaining geospatial databases at the cadastral level for the state. It has also been tasked with building awareness of geospatial data among other government agencies and the people of Kerala.
- **KILA.** The Kerala Institute of Local Administration is working to develop local government data partnerships in all districts. At the time of the diagnostic, KILA was working to create a summary of the state of data in Kerala, and a related strategy, and intends to complete this by the end of 2023. In partnership with a development agency, KILA is working on an operational manual for the Urban Observatory. Kila also plans to deliver training on data-driven decision making but does not have the resources to do so yet.

- **Department of Economics and Statistics**. The Department of Economics and Statistics (EcoStat) has 250 staff. Several of the staff are embedded in other agencies within the government, such as the agriculture or health department, to support data production. EcoStat makes some of its data public in machine-readable format even though it has not received demands for data from non-state actors.
- **LSGD.** The Local Self-Government Department is working to develop a citizen delivery platform that will have the capability to integrate data to support data-driven decision making and provision of government services.[109]
- **WRD.** The Water Resource Department has developed and maintains the Water Resource Information System (KWRIS), which includes data on water availability and demand by building upon OpenStreetMap and Google Maps. The Department collects certain real-time data through its 150 stations. Department officials consulted for the diagnostic stated that they use the Information System for analysis and decision making. The Department makes weather data public through its portal and provides full access to its information system to KSDMA. The data collected by the department is largely driven by internal needs. There appears to be demand from non-state actors as the department sells its data to academia.
- **Land Revenue, Registration and Survey.** The survey section maintains geospatial information, and has digitized almost half of all the land data in the state. The agency intends to integrate spatial data among the three departments (revenue, registration and survey) when digitization is complete (expected by 2027). The Revenue Land Information System (ReLIS) is being used to maintain data on over 1,600 villages in the state. The information in ReLIS cannot be accessed by anyone but the relevant landowner. Officials acknowledged that there could be multiple use cases if this data were public. For example, with these data it would be possible to build platforms like Zillow or Redfin.[110] This would allow people to confirm the owner and status of a piece of land before purchasing it to avoid fake documents/fraud/cheating in the purchasing process. The agency also maintains "Bhu-Naksha", a public platform that provides a land parcel map and geospatial data coordinated in a PDF file. The agency shares some of this data with other government departments via an API for which a one-time "formal paper train" approval is needed. The agriculture department appears to be the primary user of the data.
- Some state institutions have national counterparts that influence their priorities, infrastructure, resources, and relationships with other state agencies. For example, KSDMA is the state arm of the NDMA (National Disaster Management Authority), which is the apex body for disaster management in the country and lays down the key policies, plans, and guidelines for India. KSDMA can and does leverage data and other digital resources provided by the NDMA; the arrangement is however not without its complications. As discussed earlier, there is currently no agency with a clear mandate for overall data leadership. **Land Revenue, Registration and Survey.** The survey section maintains geospatial information, and has digitized almost half of all the land data in the state. The agency intends to integrate spatial data among the three departments (revenue, registration and survey) when digitization is complete (expected by 2027). The Revenue Land Information System (ReLIS) is being used to maintain data on over 1,600 villages in the state. The information in ReLIS cannot be accessed by anyone but the relevant landowner. Officials acknowledged that there could be multiple use cases if this data were public. For example, with these data it would be possible to build platforms like Zillow or Redfin.[111] This would allow people to confirm the owner and status of a piece of land before purchasing it to avoid fake documents/fraud/cheating in the purchasing

[109] PPT shared by Jobin John, Urban Domain Expert & Municipal Secretary.
[110] These are online platforms in the US that provide real estate information.
[111] These are online platforms in the US that provide real estate information.

process. The agency also maintains "Bhu-Naksha", a public platform that provides a land parcel map and geospatial data coordinated in a PDF file. The agency shares some of this data with other government departments via an API for which a one-time "formal paper train" approval is needed. The agriculture department appears to be the primary user of the data.

Government data management

The purpose of this section is to provide an overview of the data housed in various applications used by different departments and to facilitate coordinated data management, thereby promoting data sharing and decreasing redundancy in data collection. This list of application is not comprehensive, but rather a work in progress to be updated as different departments develop and deploy applications. The list was created based on in-person discussions with various stakeholders from the line departments.

Unified Registry

USER: KSITM DEVELOPED BY: KSITM

The Unified Registry has the potential to become the foundation for data governance in the state. It would:

- Underpin the Digital Kerala Architecture and become a unified government service interface for citizens by adhering to the "ask once only" principle
- Serve as a reference resource, and facilitate developing, updating, and harmonizing data standards

However, the Unified Registry is still a largely conceptual application. The design and architecture are yet to be finalized.

PWD4U

USER: PWD DEVELOPED BY: CDIT

The "PWD4U" mobile app was created so that the public can submit complaints regarding public works. The app has been used by citizens extensively and is used to crowdsource information on the status of roads across the state.

Direct Benefit Transfer

USER: Farmers and other Agri Dept stakeholders DEVELOPED BY: NIC

Direct Benefit Transfer aims to transfer subsidies directly to people through their bank accounts:

- Crediting subsidies directly into bank accounts reduces leakages and delays in the disbursement of all government transactions
- Direct Benefit Transfer has data about farmers and their banks

Kerala Open Data

USER: Public | DEVELOPED BY: NIC

The Kerala open data website, launched in 2019, is the state's primary open data repository. The website:

- Features almost 10,000 data resources from over 30 different government agencies
- Is an instance of the national open data portal (data.gov.in)

The site has the potential to become a very valuable resource for the data community in the state, but it needs to updated on a regular basis. Most datasets have not been updated since 2020 and the number of views and downloads is small. Currently it feeds data to the national open data portal by the Government of India.

eSavanam

USER: GoK | DEVELOPED BY: CDIT

The eSavanam website (and the accompanying mobile app mSevanana) provide a consolidated portal to approximately 800+ services offered by more than 80 state agencies. It had an impressive start during the pandemic when most government services moved online and citizens sometimes struggled to locate government services. However, there are a few challenges:

- It appears that the services do not follow a common data governance strategy
- Each agency collects data individually and there appears to be a lack of adequate coordination with other agencies to manage the data
- Privacy rules governing the service are very rudimentary, with no discussion of data sharing, for instance

eSevanam 2.0 is being strategized with many additional features.

LIMS - RELIS

USER: Landowners and Land Dept | DEVELOPED BY: NIC

The RELIS Land Information Management System has information including:

- Information on 1,666 villages
- Land ownership data
- Type of land data

Currently, RELIS data can only be accessed by landowners and the Land Department. Sharing this information before a land sale could help make the process more transparent.

Kerala Geo Portal

USER: GoK | DEVELOPED BY: ICFOSS

Spatial data is one of the primary focus areas of the state's data program. The GoK has launched a state spatial data infrastructure program to "share and explore data related to political and administrative boundaries, natural resources, transportation and infrastructure, demography, Agro and social economy etc. of the state". The associated Geo Portal:

- Includes State and Administrative boundaries, and currently provides almost 300 data layers, with plans to add more soon
- Consolidates and harmonizes spatial data from multiple state data agencies including the LSGD, Kerala State Land Use Board, and other agencies

While the data is open in the sense that users can download individual layers as PDF, PNG, or JPG files, it is not easy to reuse or share as there is no API and frequently the metadata does not include licensing information for reuse.

Public Distribution System

USER: GoK | DEVELOPED BY: NIC

The Public Distribution System (PDS) is one of the key sources of citizen data, especially beneficiaries of the ration card. The PDS carries out functions such as:

- Control of rationing and marketing of essential commodities
- Consumer affairs
- Consumer Dispute Redressal Commission (CDRC) and Fora (CDRFs)
- Distribution of kerosene

GRAMAM

USER: GoK | DEVELOPED BY: KSREC

GRAMAM, built on OpenStreetMap, is a grassroots mapping application (citizen mapping level capability). It is used to map multiple assets by partnering with locals/citizens/volunteers. All geospatial data to be used by/for KSMART will be mapped by the KSREC using the GRAMAM app, and other agencies are using it as well. Every month, KSREC organizes training for academics and volunteers. For example, there are 5 volunteers per ward who are contributing to map roads for the LSGD department. Thus far:

- 1 lakh km of roads have been mapped by the LSGD
- 7.5 lakh wells have been mapped by the Groundwater Department
- Wildlife and building assets have been mapped by the Forest Department
- KSDMA used the Damage Mapper Module to view and update maps during the floods

Satellite Data Program

USER: GoK | DEVELOPED BY: KSREC

KSREC, with the support of ISRO-based satellite data, is to conduct spatial depiction of land and water resources along with their attribute information for the preparation of the District Resource Geospatial Atlas, keeping Village Cadastral data as the base in a virtual, seamless manner. Related to the program, KSREC:

- Has access to 28cm level hi-resolution satellite images
- Aims at generating a geo-spatial information catalog at a detailed scale (1:10,000) for use
- Intends seamless integration with data/information at the stakeholders' end in district- and local-level planning for the entire country

Integrated Agriculture Management Information System

USER: Farmers and other Agri Dept stakeholders | DEVELOPED BY: NIC

The Integrated Agriculture Management Information System (IAMIS) has data on:

- Farmers
- Land
- Government schemes for farmers and government-affiliated vendors are mapped in

The system could benefit from integration with land revenue, KSDMA and other data sources.

Kerala Water Resource Information System

USER: GoK | DEVELOPED BY: Nabcons

The Kerala Water Resources Information System is designed to be a "single, authoritative" source of all weather-related data in the state (though interviews suggest that the system does not currently incorporate LSGD data). The system aims to be a source of insight and support for decision making, instead of a mere data repository.

- 84 Kerala state stations and 150 Kerala real-time stations are live today
- Ground water data is provided by 800+ peso gauges
- WRD has the panchayat-level boundary data from the Land Use Department, last updated in 2018/19

Package for Effective Administration of Registration Laws

USER: GoK | DEVELOPED BY: NIC

The PEARL Suite (Package for Effective Administration of Registration Laws) is a package of web-based applications that are hosted at the State Data Centre and have been implemented in 41 SROs of the Trivandrum District in Kerala to collect land ownership data, type of land owned, and other land-related data that will eventually reside in the LIMS. It has datasets gathered via:

- Stamp duty collection
- Encumbrance certificates
- Marriage certificate data

E-Health

USER: Health Dept. | DEVELOPED BY: E-Health

The E-Health team has 30+ software developers, 2 data analysts, 5 to 6 testers, 2 DBA, and 2 network administrators on staff. For data analytics, they are using tools like Power BI, Tableau, and others. The department is also in the process of creating a draft health data sharing policy.

KSMART

USER: LSGD | DEVELOPED BY: IKM

KSMART is the e-government platform working across 87 municipalities, 7 corporations and 941 Gram panchayats. Key information stored in the system includes:

- Data of 93 local bodies
- License data and citizen data
- LSGD asset data – such as roads, buildings, and manpower
- Sectoral budgets, HR/staff data, and supporting documents

IKM is in the process of integrating 14 applications. In the initial phase, birth/death registrations, trade licenses, and public grievance redressal applications will be available through KSMART. All other services, including building permit applications, are expected by the end of the year (2023-24), when the application will be launched across all urban bodies. The services will initially be launched in the Kochi and Kannur Corporations and the Attingal, Thathamangalam, Thodupuzha, Cherthala, Kodungalloor, and Anthur municipalities.

Kerala Warning Crisis and Hazards Management System (KaWaCHaM)

USER: GoK DEVELOPED BY: KSDMA

KaWaCHaM is an integrated early warning system to be used except for disasters that require mobilization of resources at the national level. The system:

- Has access to backend geo-database of over 280 layers
- Takes inputs from KWRIS and IMD for weather monitoring and prediction; 100 weather stations across the state provide data to the KaWaCHaM
- Is available to citizens via the KSDMA portal
- Has an Intelligent Decision Support System with geodata and warning dissemination – using principles of AI/ML it learns from past events to predict better
- Publishes water reservoir level data every day at 11am, using data provided by the KWRIS

One Health Initiative

USER: Health Dept. DEVELOPED BY: Health Dept.

The One Health Initiative (OHI) can analyze relationships between livestock-based data, public health, zoonotic diseases, and other variables. The initiative, encouraged by multiple relevant government departments, was prompted by the need for better information given that in the last few years Kerala has been affected by the Nipa Virus, Avian Flu, and COVID-19. In its second year, the OHI:

- Aims to train about 2.5 lakh volunteers; KILA has been engaged to conduct local level trainings and socializing workshops, since the OHI is a new concept for most
- Is focused on 4 of 17 districts in Kerala
- Uses the Integrated Disease Surveillance Program (IDSP) as one of its current data sources; data is available in Excel file format

Road Maintenance Management System

USER: PWD DEVELOPED BY: TRL Software

The Road Maintenance Management System (TRL VK iRoads) is the current RMMS. It is a web-based robust system managed and maintained by TRL. The system:

- Currently captures 155 attributes related to structural and engineering data points of the road
- Covers the 7,000 km core road network used for goods transport, out of the 32,000 km of roads under the PWD; one time data collection on the 7,000 km portion was done in 2021-22, and based on this data, Forward Work Plans were generated from the system and submitted to the government for budget allocation

Ongoing maintenance and update of this system will require innovative tech interventions.

Sanchaya for Tax Collection

USER: LSGD DEVELOPED BY: IKM

Sanchaya is an e-government application for the Revenue and License System in local governments. Data points collected and services provided by Sanchaya portal include:

- Know Your Property Tax (Sanchaya Tax)
- Online payment of property taxes and levies
- Application for Ownership Certificate
- Application for Building Age Certificate
- Rent to civic bodies
- E-filing
- Dangerous and Offensive
- Property Tax Search using Plinth Area
- Utility bill payment (water, electricity, or telephone)
- Profession Tax, Hall Booking, Entertainment Tax (Cinema)

Hospital Management System

USER: PHC, Medical colleges DEVELOPED BY: E-Health

Ayushman Bharat Digital Mission (ABDM) is an initiative by the Government of India to digitize public health records. They have created a milestone-based system for all Hospital Management Systems (HMS) to communicate with each other via the Ayushman Barat platform, as follows:

- Milestone 1 – Creation of an ABHA ID for the patient
- Milestone 2 – Make the stored data shareable in predefined formats
- Milestone 3 – Integration with the ABDM platform

E-Health Kerala is on Milestone 1, with an HMS that is used by almost 500 medical institutions (government hospitals and medical college hospitals) out of 1,300 institutions. It has:

- Patient data records
- ABHA ID (Ayushman Bharat)
- An FHIR-based data sharing mechanism (Fast Health Infrastructure Resources [FHIR] has defined standards for data sharing across HMS globally)

The plan is to roll out the HMS to all government hospitals in Kerala eventually.

Given the macro-focus and rapid nature of the data diagnostic, the diagnostic team has not been able to conduct detailed heuristics on the individual applications or gather structured feedback from their users and administrators. The interviews suggest that most services work reasonably well and are effective at meeting their core functions. Several challenges were observed, however:

- There is currently no central data inventory that can show what kind of data the government produces, in which format, how often it is updated, and who owns it. Government employees appear not to know who to ask for what. As described earlier, data sharing is currently personality driven and ad hoc. This prevents the government from effectively deploying its resources for data production (interviewees raised concerns about the risks of producing the same or similar data more than once). Uncertainty about the "golden" source of data also means that even when data exists, government agencies are not able to fully leverage it for development.
- Difficulty acquiring data from national platforms – for example, the national databases (India disaster resource network, NDMA, NDRF, DCRA, NDEM, NIDM, MNREGA) that feed into the state disaster risk fund are not easy to link to in the absence of an API or real-time access.
- Difficulty developing granular local data – there are many initiatives currently underway to capture a variety of local, granular data but there is no structured state level data management strategy to unify these initiatives and help different agencies pool and/or share data.
- The data sharing practices within government are inconsistent:
    - There is a lack of clarity about appropriate policy guidance for sharing data among different government agencies.
    - There is a lack of clarity about the ownership of different data variables.
    - There is no standard data production mechanism. Hence, when data is produced, it is likely to be on an ad-hoc basis, and published without metadata, and in a way that limits its use beyond its initial purpose.
- The lack of digitization is a constant theme:
    - Much water-related data has not yet been digitized.
    - The tax register is in the process of being digitized.
    - A digital survey of villages is ongoing.
- Constant debate about what data can and should be open. (Resistance to publication of risk-related data that might lead to rezoning or otherwise affect land prices.)
- Many agencies do not appear to understand that data must be curated and sanitized after it is collected.
- Need for last-mile data collectors.
- Government departments are no longer stand-alone; most activities are now inter-departmental, but the systems are still departmental – need for a data curator to combine activities across departments.
- The state data center (tier 3) is commonly used. As a future exercise the GoK can consider whether an upgrade to tier 4 would be appropriate.

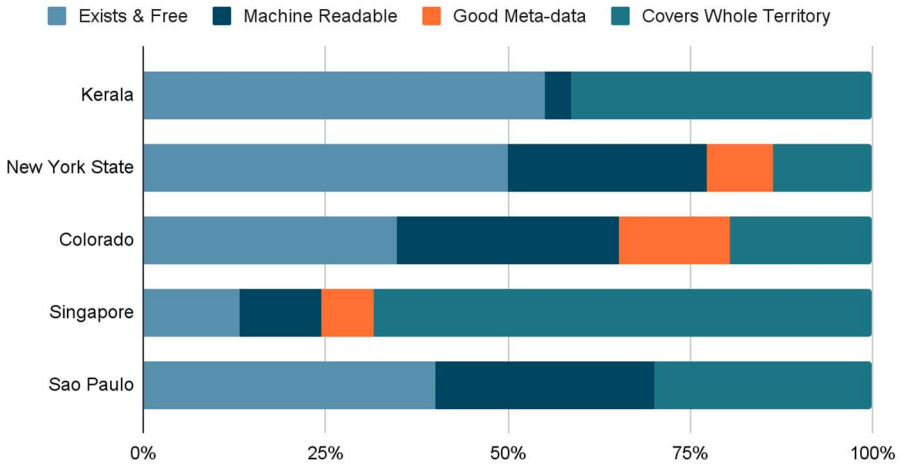
STATE OF DATA OPENNESS IN KERALA

The diagnostic team compared Kerala's data openness against four other national and subnational jurisdictions (the states of New York and Colorado in the US; Singapore; and the city of Sao Paulo in Brazil)

using an evaluation matrix[112] that considers the availability and quality of critical disaster-related data. Overall, Kerala ranks well compared to the rest and has the greatest amount of disaster-related data available at no cost to the public (Figure A1). However, there is room for improvement. The GoK could provide more data in machine-readable formats (such as geospatial data in the form of shape files instead of PDFs) that will aid analytical work. Metadata can also be improved. By having updated explanatory notes regarding each dataset, greater data consistency and compliance can be achieved.

Lastly, Kerala's data currently resides across at least four different websites, which makes it difficult to search for information. The comparator states have consolidated data repository websites, having information all in one place and reducing duplication across agencies.

Figure A1. Comparison of Data Openness in Kerala with select other States



Source: World Bank Staff

Demand for data

Even as the state has developed several digital platforms and services, very few agencies appear to work closely with stakeholders to assess the demand for and the effectiveness of the solutions they have developed. Consultations indicated the following:

- The GoK currently does not systematically measure or assess the demand for data. Interviews suggest that administrative boundaries and survey parcels are the most requested spatial data, but this is difficult to verify in the absence of regular measurement and documentation of demand.
- The GoK sets no performance targets for data use, which may also explain why the lack of demand is not a priority in many agencies.
- The state's open data portal has been dormant since its launch and the datasets, which have not been updated since launch, are barely accessed or used.
- Some state agencies take more interest than others in growing the demand for their data, and the value-add they provide by cleaning and curating raw data makes their data appealing to other

[112] The evaluation matrix was developed by the team to analyze and compare the availability and quality of critical disaster-related open data across the few chosen jurisdictions in a structured and systematic way. Data important to disaster-risk management (such as Flood Hazard Maps, and Administrative Boundaries) were assessed against critical factors (such as availability of the data, format of the data, and cost of obtaining the data) and attributed with either a 'Yes' or 'No'. The final scores were tabulated based on the percentage of the number of 'Yes' within each jurisdiction against the total number of factors and compared across.

agencies within government. KSDMA for instance provides important geospatial data to the WRD, among others. KSDI is also gradually emerging as a platform gaining wider recognition across the state.

- Many state agencies make it difficult for non-government actors to use their data:
  - KSDI for instance does not provide an API, and its data while technically open is only available in "dumb" formats like PDF that machines cannot easily read or consume.
  - Private entities and academics must also go through an approval process before KSDI shares data with them in usable formats. There are no published criteria currently available to explain why certain requests are approved or denied.
  - There is often no licensing information available about the datasets that the government releases. This makes it risky in some cases for non-government entities to develop commercial applications that rely on such data.
- There does appear to be appetite to source data from commercial sources, sometimes even when another state agency already produces comparable data. Many government agencies for example continue to buy spatial data from Google even though similar data is available for free from KSDI.
- To distribute their apps, the creators of the apps have used channels like Apple Inc.'s App Store and Google Inc.'s Google Play. Ratings and reviews of the apps on these distribution platforms offer crowdsourced insights to potential new users about the app quality, which likely impacts uptake.[113] For this diagnostic, a quick scan of available apps on Google Play surfaces Pol-App,[114] a police app that has been downloaded over 500,000 times (See Table A1). This app, which can be used by residents to inform the police when they travel out of town, appears to be popular.
- The eSevanam[115] website is said to attract 75,000 visitors.

Table A1. Select Apps related to Kerala public services available on Google Play

No.	App Name	Offered by	Minimum download	No. of Ratings	Score (Out of 5)
1	Kerala Tourism	Kerala Tourism Official	100,000	NA	NA
2	Pol-App (Kerala Police)	Kerala Police	500,000+	6,000	3.5
3	Kerala Pension	Department of Treasuries, Government of Kerala and NIC-Kerala	100,000+	1,000+	2.0
4	Digital Diary	IPRD, Govt of Kerala	1,000+	NA	NA

Source: World Bank Staff using Google Play data as of May 2023

- One reason for a generally low level of demand for data within government may be the limited capacity of state agencies to use data effectively. Many agencies find difficulties in conceptualizing new data-driven solutions or integrated data-driven processes within their current service delivery models.

[113] Rajesh Vasa, Leonard Hoon, Kon Mouzakis, and Akihiro Noguchi. 2012. "OzCHI '12: Proceedings of the 24th Australian Computer-Human Interaction Conference." November 2012, pages 241–244. Available at https://doi.org/10.1145/2414536.2414577.
[114] Available at https://play.google.com/store/apps/details?id=com.keralapolice&hl=en_IN.
[115] See https://services.kerala.gov.in/.

Civic engagement

Civic engagement is a vital component in the success of any data initiative. When data is used and reused by a broad group of users, it can help improve the quality of data. Mutual connections and understanding between government agencies and civil society are important to collectively learn and explore how to do open data well. Open data results in new communication and interaction channels. To realize open data's value, it is important that civil society and private sector stakeholders serve as willing and capable intermediaries, preparing data for broader usage, and creating applications and visualizations that disseminate information and insights to the general public. Some findings and experiences that emerged from the diagnostic include:

- Kerala Startup Mission (KSUM) is the GoK's nodal agency for promoting entrepreneurship in the state and supporting the startup ecosystem through various technology support programs. Since its inception in 2006, KSUM has supported 3,900+ registered startups, disbursed over 30 Crore in grants, and supported 63+ incubators, 23 fab labs,[116] a digital fabrication laboratory, and 300+ innovation centers across the State. KSUM's successful operation is in large part due to its engagement with the private sector, students, civic society, and even Non-Residential Indians. KSUM runs specific programs that bring in social entrepreneurs from rural and marginalized communities. Problem cells created within government departments help articulate needs, and KSUM connects these needs with tech solutions (through a demand-based as well as supply-based system). KSUM has also organized hackathons, including on the topic of resilience.
- KSDMA has a history of providing timely public notifications during emergency events. The department has recently developed KaWaCHaM (Kerala Warnings, Crisis and Hazards Management system) with analytical capabilities on disaster and climate change risk. A component of KaWaCHaM is planned to be public facing. Currently it does not work for Red Level warnings.
- The PWD4You app crowdsources user-generated data, and is used by the PWD for public complaints and the redressal thereof. IKM also has an online public grievance mechanism, as well as 16 other digital solutions for citizen services; however, citizen consultation was not incorporated into the design of these solutions.
- KSDI has organized multiple mapathons (based on the Open Street Map [OSM] platform) to expand the collection of hyper-local mapping data and engage community volunteers to do so. The mapping is performed by universities, students, and other civic volunteers.
- KILA often helps with public consultations, especially on master planning, but it has not held any public consultations on data sharing between government and civil society. Some civic engagement has taken place through academia.
- KSREC is the state counterpart to the Indian Space Research Organization (ISRO), the nodal agency for generating geospatial data on land use (updated every five years). Managing wasteland and land degradation are two of the key tasks undertaken by KSREC. They depend on satellite, drone, and at times aerial imagery. GRAMAM, built on OSM, is an app created by KSREC for mapping roads by partnering with locals. There are five volunteers per ward contributing to the GRAMAM initiative. KSREC also allocates 5 percent of its budget to training, and organizes training for academics and volunteers every month.
- The idea of citizen-developed apps, as well as community-led data gathering (participatory data collection owned by the community, with bias correction) and maintenance of IoT sensors, has great potential. These initiatives could be useful to inform citizen decisions in real time about their vulnerabilities and options.

[116] See: https://sites.google.com/startupmission.in/fablabkerala/home_1?authuser=0

When it comes to citizen engagement, the GoK mainly focuses on public communication portals where public outreach is sought, and participation is invited. It has not yet extended these outreach efforts to actively promoting the use of government data or involving the community in data gathering and maintenance of sensors/resilience monitoring.

Funding

Overall, agencies consulted for this diagnostic indicated they do not have reliable funding for their long-term data initiatives. There are proposals to allocate 2 to 3 percent of the state budget to IT-related activities, with the expectation that this funding would include support for data initiatives.

Officials expressed their concern that there is usually no dedicated team to work on funding schemes to sustain data initiatives over the long run. The lack of adequate funding combined with the shortage of officials with relevant data skills creates a significant barrier to the state's aspirations around using data to strengthen resilience.

In an agency like KSREC, there is dedicated funding to support data initiatives. KSREC also has the flexibility to partner with academia and the private sector to bring in funding to support some of its data initiatives. Notably, KSREC has allocated part of its budget to provide trainings to a diverse group of stakeholders.

Based on the GoK's 2020-21 budget announcement, the government is also working to create a public-private partnership to invest in data science and artificial intelligence.[117] The government plans to set up a new IT facility with an area of 500,000 sq ft. Additionally, Rs 50 Crore has been earmarked to help "educated housewives" get IT jobs "near home", and 2,000 Wi-Fi hot spots are to be established to help citizens access government services. The government also plans to invest in "science parks", a "hardware technology hub" and a "green mobility technology hub".[118]

However, given a general lack of clarity on how the government allocates budget to enable broad-based data use, the diagnostic team recommends creating a predictable and sustainable source of funding to support the data for resilience agenda.

Technology and skills infrastructure

Through a series of initiatives, including the Kerala State IT Mission and Kerala Digital University, the GoK has invested in the enabling environment for data skills development. For example, the Digital University provides many relevant programs for building the skills of government employees.

Generally, the GoK recognizes the importance of and invests in training[119], study tours[120], and post-graduate courses[121] on e-governance for GoK employees. The spatial data team at the KSITM conducts GIS training for other state agencies at no cost to them. This is especially important as most state agencies that supply data to the KSITM do not have permanent or dedicated GIS teams and thus have a constant need for retraining. The team also organizes periodic training for student volunteers who participate in the mapathons organized by the team. Training is designed to ensure that students understand basic GIS concepts and collect data in a standardized fashion.

---

[117] See https://www.finance.kerala.gov.in/includeWeb/fileViewer.jsp?hDx=budget!2020-21!Speech2020-21-Eng.htm.
[118] See https://economictimes.indiatimes.com/news/economy/policy/boost-to-it-industries-infrastructure-in-keralas-first-paperless-budget/articleshow/90153917.cms?from=mdr.
[119] See http://elearning.kerala.gov.in/.
[120] See https://timesofindia.indiatimes.com/city/thiruvananthapuram/kerala-chief-secretary-led-team-visits-gujarat-to-study-e-governance/articleshow/91138379.cms.
[121] See https://rdd.lsgkerala.gov.in/system/files/2021-03/PGDeG%202021-22.pdf.

There are also initiatives being discussed to further strengthen the government talent pool. For example, the government is currently exploring the possibility of supporting employees so they can pursue a two- to three-year university course.

To get a preliminary perspective on the local availability of relevant talent, the diagnostic team scraped data from LinkedIn using "data" as a keyword for jobs and setting job search location to "Kerala, India". The platform returned search results of over 5,000 related job postings, and over 126,000 people who have data-related jobs. A sample of around 800 results with job postings dated from Feb. 12, 2022 to Dec. 4, 2022 was used to gain quick insights into the data job market in Kerala. The top five cities in Kerala with data job vacancies were Kochi, Thiruvananthapuram, Kozhikode, Thrissur and Ernakulam respectively (see Figure A2). Kochi alone accounted for 47 percent of the job vacancies in the sample size.

Of the data jobs available, mid- to senior-level executives (categorized as experienced professionals with 5-10 years of experience) were most in demand. The engineering and IT industries were hiring the most, accounting for 36 percent of the job vacancies available (see Figure A3).

Figure A2. Cities with the most data-related job vacancies in Kerala

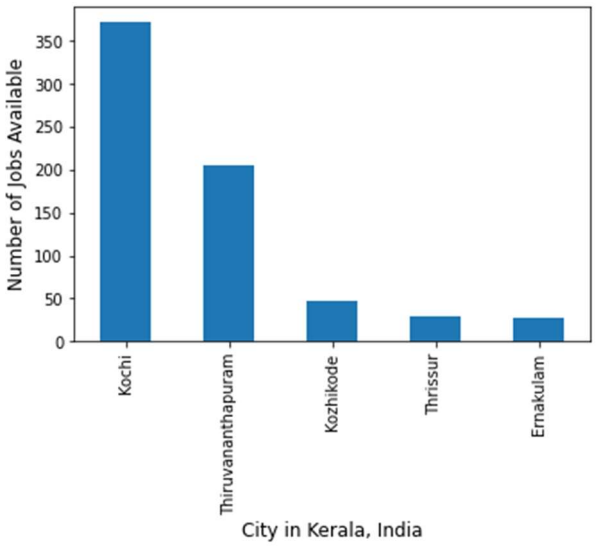
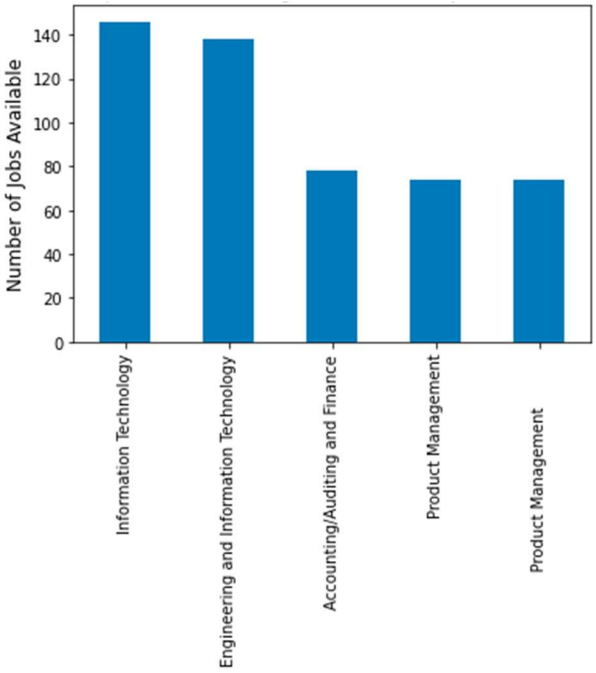


Figure A3. Industries with the most data-related job vacancies in Kerala



Source: World Bank staff analysis of data from LinkedIn.com

Appendix 2. List of interviewees

Table A2. In-person consultations conducted in Kerala for the data diagnostic during January 2023[122]

No.	Department	Designation
1	KSITM (Kerala State IT Mission)	Director
		Spatial Data Officer, Kerala State Spatial Data Infrastructure
		District Project Manager (HQ)
		Mission Coordinator, Unified Registry
2	Rebuild Kerala Initiative	Deputy CEO, RKI
		Program Executive (Social), RKI-PMSS (currently not in position due to contract expiration)
3	Electronics & IT Department	Secretary
4	Irrigation Design and Research Board (IDRB)	Assistant Director
		Assistant Director
5	Irrigation Department	Chief Engineer
		Joint Director
		Assistant Exe. Engineer
		Deputy Director
6	Survey & Land Records Department	Deputy Director
		Project Manager
7	Dept of Land Revenue	Assistant Commissioner Nodal Officer IT
		State Coordinator IT Cell
8	LSGD Planning	Deputy Town Planner
		Town Planner
9	Information Kerala Mission	Urban Domain Expert
10	Dept of Registration	Joint IG of Registration
		Nodal Officer IT
11	KSDMA	Member Secretary KSDMA
		GIS Specialist
12	KSREC	Scientist
13	PWD	Executive Engineer
14	KILA	Urban Chair Professor
15	Department of Economics and Statistics (DES)	Director
		Additional Director
16	Principal Secretary Sports and Youth Affairs	Ex Principal Secretary IT (now retired from service)
17	eHealth Kerala	Manager Technology
		Information Security Officer
18	One Health Platform	Public Health Specialist & Assistant Director (Medical), Directorate of Health Services
19	Corporation of Trivandrum	Secretary Trivandrum municipal corporation
20	Kerala StartUp Mission	Senior Manager, startup life cycle & IT

[122] The WB Diagnostic Mission team met these officials in January 2023. The information collected was aggregated and analyzed to present overall trends and does not necessarily reflect view of a particular official or department.

Appendix 3. Diagnostic questionnaire

This diagnostic was initiated as a response to the GoK's focus on open data, one of the cross-cutting pillars in the framework for the government's Rebuild Kerala Initiative.[123] The diagnostic team adapted ODRA (Version 3.1), prepared by the WB's Open Government Data Working Group, to prepare a questionnaire for stakeholders that reflected the cross-cutting nature of the resilience lens. The ODRA assists in identifying actions that a government authority could consider to establish a data program either at the sub-national level, or for an individual public agency. It consists of a rapid diagnostic of the eight dimensions considered essential for the success of a data program. Key questions for dimensions below.

SENIOR LEADERSHIP: Data programs require the implementation of changes – often including legal, institutional, technological and cultural changes – and may affect stakeholders both inside and outside government. Focused, strong, sustained, political/senior leadership is therefore critical to helping a government overcome resistance and inertia of all kinds, to helping incentivize actors to make the necessary changes in a timely and effective manner and to achieving the desired objectives and benefits.

A. Has the GoK articulated a vision for data-led digital transformation of the state (or any priority sectors, including resilience)?
    a. Please provide links to announcements or initiatives demonstrating the government's commitment to data-driven development (especially for resilience).
    b. Please include references to any roadmaps or performance targets included in the government's vision of digital transformation.
B. Which government agency/official is responsible for coordinating data policy and implementation activities in the state, and coordinating them with inter-state or national initiatives?
C. Please list local Civil Society Organizations/NGOs whose leadership has made public commitments to using data/digital tools to increase resilience in the state (provide links when possible).
D. Please list local private sector business leaders closely associated with developing or using data-driven tools for resilience.
E. What socialization around data has been done at the political level? What awareness-raising events/activities around data have been held for agencies or the general public, and what political or senior leadership involvement was there?
F. To what extent do current political priorities align with different motivations for using data: strengthening transparency and accountability; economic growth; inclusion and empowerment; public services; and government efficiency?
G. Where is Kerala State in the political cycle? What scope is there for sustained momentum to release data before the next election?
H. Which political priorities could be significantly furthered by data?

POLICY/LEGAL FRAMEWORK[124]: The long-term success and sustainability of a data program depends greatly on the enabling policy and legal framework. Leveraging data effectively requires that a range of policy and legal issues be addressed – for example, with respect to the licensing and reuse of data, ensuring privacy and data protection, and anonymizing personal and personally identifiable data. It is important to identify at an early stage the existing policies, laws and regulations with respect to a core set of issues, and to identify actual or perceived obstacles in order that essential policy or legal changes can be initiated early. It is recommended that qualified local counsel, familiar with these subject matter areas, assisted by qualified "international" legal counsel with relevant experience, perform the legal diagnostic.

[123] GoK framework for Rebuild Kerala Initiative.

[124] The preliminary analysis and recommendations in this section are based on information and opinions collected from interviews undertaken and materials provided by the government and other local stakeholders during this study. This section is not based on detailed, legal due diligence and does not constitute legal advice. Accordingly, no inference should be drawn as to the completeness, adequacy, accuracy or suitability of the underlying Diagnostic, or recommendations, or any actions that might be undertaken resulting therefrom, regarding the enabling policy, legal or regulatory framework for Data in the Kerala. It is therefore recommended that, prior to undertaking any legal action to address any legal Diagnostic issue raised herein, a formal legal due diligence be performed by competent, locally qualified legal counsel, preferably assisted by international legal experts with relevant experience and knowledge of these areas.

A. What activities or programs has the government initiated to improve (for each, please provide links and references, or supporting documents):
   a. the data infrastructure in the state, including data/computing platforms that provide hardware and software services to businesses and CSOs, apart from government agencies?
   b. the state-level regulatory environment for the collection, sharing, and use/reuse of data (online, through IoT devices including sensors and drones, and for tools that utilize AI)?
   c. the development of digital literacy and capacity building in the state, including that of GoK officials?
   d. promotion of data innovation and entrepreneurship in the state?
   e. mitigation of harms from the misuse of data (especially in the context of marginalized and otherwise vulnerable populations) within the state?
B. Which agencies officially sell their data? To what extent are any such agencies dependent on revenues from sale of data for their annual budgets?
C. What exclusive arrangements does government have with companies with respect to any datasets?
D. Has the GoK provided any guidance on the use of emerging data-driven technologies such as AI, IoT, and blockchain? Does the GoK provide any guidance on algorithmic-decision systems, and the use of automated or semi-automated systems by the state government?
E. What official policies, laws or rules require the anonymization of data prior to publication?
F. Does the GoK have any additional requirements that exceed provisions stipulated in national regulations?
G. What official policies, laws or regulations address data security, data archiving and digital preservation?
H. What is the government's experience to date with data security?
I. How does the government (or any agency or local authority) license or permit the release/use of its data?
J. Has the government established a policy framework for the industrial internet (or Industry 4.0/IoT); what are the key provisions for data collection using sensors, consent requirements for machine data, and data sharing between owners and suppliers of equipment?
K. Has the government established a policy framework for digital capacity building and digital inclusion (especially of marginalized communities); are there any provisions regarding school curricula, after-school or work capacity building programs, or requirements for local language delivery of online services?

INSTITUTIONAL STRUCTURES, RESPONSIBILITIES AND CAPABILITIES WITHIN GOVERNMENT: In addition to political and senior leadership, middle management level skills and leadership are important to success. Creating a data program requires agencies to manage their data assets with a transparent, organized process for data gathering, security, quality control and release. To effectively carry out these responsibilities, agencies need to have (or develop) clear business processes for data management as well as staff with adequate IT skills and technical understanding of data (e.g., formats, metadata, APIs, databases). Engagement among agencies and at all levels of government to set common standards and remove impediments to data interoperability and exchange is also vital, and requires mechanisms for inter-agency collaboration. In addition to handling the "supply side" of creating a data program, agencies need the structures and capabilities to engage with communities that reuse data – including developers, companies, non-governmental organizations, other agencies and individual citizens.

A. Provide a document/organization chart describing how data/digital institutions are organized within the GoK. Identify nodal/coordinating agencies and indicate if any include CIO- or CDO-type officials.
   a. Which of these agencies are engaged in resilience-building activities?
   b. Have any of the agencies above been assessed for relevant capabilities, project management experience, and technical skills? Please share relevant documentation if available.
   c. Has any agency been granted the formal mandate to be the lead institution in the planning and implementation of a data program?
B. What inter-agency mechanisms are used to coordinate IT issues (such as for technical matters)?
C. What process is currently used to measure agency performance or quality of service delivery specifically related to disaster management, rescue, relief, and rebuilding?
D. Which agencies or ministries have been the most active in the adoption of data-driven tools and platforms for service delivery, and in releasing their data? Conversely, which agencies and departments have expressed most concern about the release of data, and the adoption of digital technologies? What is the stated basis of their concern? How can they be handled procedurally, and how can their concerns be addressed?
E. Which agencies are responsible for coordination and engagement on data/digital with private sector/CSOs?

F. How strong is the GoK's overall IT skill base among senior government leaders and civil servants?
   a. To what extent are leaders "digitally literate"?
   b. To what extent do government officials receive training on IT, data standards or data analytics?
   c. What diagnostic approach does the government use (e.g., through an IT competency framework) to determine or track IT skill levels among civil servants?
   d. What part do IT skills play in deciding civil service grades and promotions?

GOVERNMENT DATA MANAGEMENT POLICIES, PROCEDURES AND DATA AVAILABILITY: Data programs can build on established digital data sources and information management procedures within government where they already exist. Where data is only available in paper form it will be hard to release it in reusable format quickly and cheaply. Good existing information management practices within government can make it much easier to find data and associated metadata and documentation, identify business ownership, assess what needs to be done to release it, and put processes in place that make the release of data a sustainable, business-as-usual, downstream process as part of day-to-day information management. The diagnostic should consider whether key datasets are available and what would need to be done to release them. Identifying which datasets are key will depend on the context in Kerala. Datasets that support political or national priorities will be of particular importance because of their utility in supporting broader change. Experience from existing open data programs has shown that certain types of prominent datasets tend to support typical open data objectives, such as economic growth, more effective public services, increased transparency, and improved government accountability and responsiveness. Table A4 presents a list of the kinds of datasets that may be of interest from a resilience perspective. Key questions for this pillar include:

A. Has the government established a whole-of-government framework for digital government and are related IT architecture and operational processes defined and established?
B. Provide links to policies and practices on the management of GoK information. These should include references to policies/guidelines on data security, data quality, data archiving and preservation, and data sharing.
C. What data inventories exist, both at the whole-of-government level and at the agency level (e.g., the IT Ministry of State Statistics Office)? What standards for these apply across the GoK as a whole?
D. Are processes in place to share core government data (business registry, land database, addresses) electronically across different government agencies?
E. In what technological formats and standards is government data held? Are they open or proprietary formats? What government interoperability framework exists, and is it actively used by agencies to support the development of integrated data assets and information exchange?
F. Does the government deploy and use big data technologies (examples include cloud infrastructure, IoT devices, large scale online data platforms, AI for services)?
G. What procurement guidelines apply when the government acquires data from third parties? Are there any standards for "ownership" (especially for data generated by contractors, and in partnership arrangements), reuse/integration, resultant IP, or sharing with other government agencies?
H. Are there requirements for public agencies to identify dataset custodians (e.g., GIS Officer) or appoint a data administrator within each ministry who is tasked with coordinating data policy?
I. What is the government experience with algorithmic decision making? What processes/tools are in place to make sure algorithms are used responsibly by the government?
J. How is data archived and digitally preserved once it has ceased to be used operationally? What are the standards, policies, responsibilities and procedures for data archiving and digital preservation? To what extent do they conform to international standards and best practice?

Table A4. Indicative datasets related to resilience

No	Title	Description	Data
1	Budget data (both at Ministry of Finance and individual agency level)	Local budget, national budget	National government budget at a high level (e.g., spending by sector, department, etc.).
2	Disaggregated public expenditure and grant data (e.g., which school got what money, when)	Transaction-level public expenditure	Records of actual (past) national government spending at a detailed transactional level. and at the level of

			month-to-month government expenditure on specific items
3	Statistical data, from National Statistics Office or departments publishing recognized national statistics	National statistics, infrastructure, wealth, skills	Key national statistics such as demographic and economic indicators (GDP, unemployment, population, etc.).
4	Census data	Census	
5	Parliamentary data including records of proceedings, draft laws under debate and enacted legislation	Government accountability and democracy: government contact points, election results, legislation and statutes, salaries (pay scales), hospitality/gifts	Results by constituency / district for all major national electoral contests; all national laws and statutes to be available online
6	Procurement data (who was awarded what) and contract data (details and documents of deal)	Contracts let, calls for tender, future tenders	
7	Data on public facilities: location and services available (schools, hospitals, police stations, public toilets, libraries, GoK offices, etc.)		
8	Public service delivery and performance data at the level of individual school, hospital/clinic, etc.	List of schools; performance of schools, digital skills; health performance and prescription data, housing, health insurance and unemployment benefits	
9	Transport data including roads and public transport	Public transport timetables, access points broadband penetration	Timetables of major government operated (or commissioned) national-level public transport services (specifically bus and train)
10	Crime data (e.g., individual crimes and their locations)	Crime statistics, safety	
11	Reports of inspections, official decisions, rulings in reusable form, e.g., public health inspections of food outlets		
12	Official registers - company, charities, cadastral/land ownership, etc.	Company/business register	List of registered (limited liability) companies including name, unique identifier and additional information such as address, registered activities
13	Geospatial information - maps, address registers, points of interest	Topography, postcodes, national maps, local maps	High level map at a scale of 1:250,000 or better (1cm = 2.5km) database of postcodes/zip codes and corresponding geospatial locations in terms of latitude and longitude
14	Weather data	Earth observation data, including meteorological/ weather, forestry, agriculture, fishing, hunting	
15	Construction data (permits, zoning)		
16	Real estate data (sales, listings, taxes, property-specific data)		
17		Pollution levels, energy consumption	Aggregate data about the emission of air pollutants, especially those potentially harmful to human health
18		Global development: aid, food security, extractives, land	
19		Science and research; genome data, research and educational activity, experiment results	

DEMAND FOR DATA AND DATA SERVICES: The value of data is in its use. A strong demand-side "pull" for data is important to create and maintain pressure on government to release data. Demand is also key to ensure that a wider data ecosystem develops and that data is turned into economically or socially valuable services for citizens. The "pull" can come from civil society, the private sector, international organizations, donors and individual citizens.

A. What local businesses or academics/universities use geospatial data and maps, or weather or transport information? What data do they think the government should make available? Please provide examples and use cases of the use/reuse of government data by third parties.
B. What is the process for identifying & meeting demand for resilience-related data, outside & inside government?
C. What businesses (local firms or branches of international firms) exist to provide value-added services for business-to-business commerce, e.g., mobile big data, business directories, data on risk factors to residences or infrastructure? What GoK data would they like to see released? Provide links to relevant businesses or firms.
D. How are government priorities for data provisioning decided, and on what evidence?
E. Describe use cases that reflect demand for data that the government is both able and unable to meet.
F. Which non-government communities use government data to develop products or stories to inform the public about climate-related or health-related risks?
G. Please describe the demand for computing/data infrastructure from CSOs and SMEs.
H. What is the known extent of actual and latent demand for data from within government, and between government entities? Please provide relevant links and references.
I. What process exists for citizens to request information from public agencies / government? What information is most requested?
J. What is the level and nature of actual demand and latent demand for data from civil society, development partners and the media?
K. How involved do external stakeholders consider they are in priority setting for key national information infrastructure data such as statistics and geospatial reference data?

CIVIC ENGAGEMENT AND CAPABILITIES FOR DATA: Experience among leading governments has demonstrated that data initiatives are more sustainable and higher impact when governments follow an "ecosystem" approach – meaning governments not only invest in supplying data, but also address the policy/legal framework, institutional readiness, capacity building (for government/infomediaries), citizen engagement, innovation financing, and technology infrastructure. Governments need to play a multi-dimensional role, creating new types of partnerships across stakeholders.

A. What programs has the government established for civic engagement and outreach related to digital/data-driven development?
B. Has the government identified, and does it work with, any intermediary organizations – CSOs, media, or others – to promote digital activities and build civic capacity to use data safely and responsibly? Please provide links to associated activities.
C. Has the GoK articulated/established processes for user-centered design of digital services? And does the design of digital services account for location, language, connectivity, gender, skills, and affordability?
D. What level of citizen-to-citizen engagement exists on key political and social issues? Is this driven by data to any extent? How would the availability of government data improve this debate?
E. Which technical schools or universities have computer science, data science or disaster risk management programs? How many graduates are there per year with technical degrees?
F. How extensive is the use of social media by citizens? What are the main platforms?
G. How do government/individual agencies use social media or other forms of digital engagement? What policies exist for this, if any?
H. To what extent have citizens digitally engaged with government? What are the main issues that generate engagement? How would citizens know if their input was considered?
I. What level of citizen-to-citizen engagement is there on key political and social issues? Is this driven by data to any extent? How would increased availability of government data improve this debate?
J. Which computer/web science academics connect with the government?
K. What relevant networks of researchers or research centers exist?
L. To what extent does the secondary school curriculum include computer science or IT training?

M. What programs address relevant tech (e.g., web science, semantic web, big data technologies)?

FUNDING A DATA PROGRAM: Funding with respect to both the "supply side" and "demand side" of data is important to ensure that the objectives of a data program are met.

A. What resources have been earmarked for funding a data program for resilience? What is the source of this funding and is it contingent on additional approvals and performance targets?
B. What already exists in terms of dedicated staff for data management, both for the overall data program and among key agencies?
C. What funding is available to build the capacity of civil society? What funding (e.g., from donors or non-government sources) is there for innovations to promote "data-enabled resilience"?
D. What public-private partnerships exist related to technology that have made investments in data-driven solutions?
E. What private sector investment has been made in data/digital initiatives in the state, especially in the context of resilience? Please provide links, even if the examples are anecdotal.
F. What funding mechanisms does the government have for innovation?
    a. What funds does the government have for development of applications or e-services?
    b. What programs does the government have to support or promote entrepreneurship, start-ups or development of SMEs?
    c. What funding is available to build the capacity of civil society? What funding (e.g., from donors or non-government sources) is there for innovations to promote "data-enabled resilience"?
    d. What public-private partnerships exist related to technology?
G. What funding is available to support the necessary IT infrastructure and ensure enough staff have the skills needed to manage a data program?
    a. How could common infrastructure across agencies be leveraged?
    b. What already exists in terms of dedicated staff for data management, both for the overall data program and among key agencies?
    c. What funding and technical skills in key agencies exist to get data supplied to a data portal (including curation and cleaning of data)?
    d. What diagnostic has been made of vendor skills and what they might cost?

TECHNOLOGY AND SKILLS INFRASTRUCTURE: In very practical ways, data programs normally rely for their success at least in part on the national technology infrastructure and the IT skills of officials, infomediaries and the general public.

A. What data/computing infrastructure has been made available to state actors, within & outside the government?
B. What is the level and cost of internet access, both by broadband and by mobile technologies?
C. To what extent do ministries or agencies outsource IT functions or services to the private sector?
D. What is the startup scene? What incubators/accelerators are there for entrepreneurs or startups? Are there any venture capital or early-stage investment funds?
E. Has the government established any digital skills programs within universities/schools? Do similar programs exist for government officials?
F. How readily available is computing and storage infrastructure?
    a. What is the availability of web hosting services?
    b. What is the availability of virtual server services?
    c. What access is there to cloud computing services including processing and storage?
G. How strong is the IT industry, developer community and overall level of digital literacy?
    a. What are the statistics on the size of the local IT industry (e.g., in terms of employment, revenues or percentage of GDP)?
    b. What does the local software development/web design industry look like? How large is it? What organized communities of developers exist?
    c. What industry associations or groups for IT companies exist?
    d. What is overall level of Internet use skills?
    e. What is overall level of data skills?