To

Chapter FOUR of Vol.1

DATA: OF, FOR, & BY THE PEOPLE

Before we go through the details, let's go through a

Quick Summary

- **Given technological advances in gathering and storage of data, society's optimal consumption of data is higher than ever.**
- **As private sector may not invest in harnessing data where it is profitable, government must intervene is creating data as a public good, especially of the poor and in social sectors of the country.**
- **Data can be created as a public good within the legal framework of data privacy.**
- **In thinking about data as a public good, care must also be taken to not impose the elite's preference of privacy on the poor, who care for a better quality of living the most.**

CHROME IAS

Now, it's time for the

Details

DATA ECONOMICS & SOCIAL WELFARE

- **People produce data about themselves and store this data on public and private servers, every day, of their own accord.**
- **Data that would have involved a laborious survey to gather a few decades ago is today accumulating online at a near-zero cost.**
- **Data is still relatively expensive to process because it tends to be noisy, heterogeneous and inconsistent across sources, but technology is incessantly developing solutions to these problems.**

- **Once processed, the cost of disseminating insights is negligible – it is nearly costless to transfer information through the internet. However, dissemination of data entails another cost – that of ensuring data privacy and security.**
- **Accidental data leaks may bring forth legal consequences and substantial financial implications.**
- **The marginal benefit of data is higher than ever. A multitude of scenarios exist in which harnessing the marginal unit of data can lead to sharp increases in public welfare. For example,**
    - **A district education officer can make better decisions if he knows, for each school in his district, attendance rates of students and teachers, average test scores and status of school toilets.**
    - **Similarly, parents can make better decisions about which school to send their children to if they know the average absenteeism rate of teachers in their village and can compare the rate to that in the neighbouring village.**
    - **An app can inform farmers of the prices of produce across the country. Data has the potential to integrate markets nationally, reduce the need for middlemen, reduce prices for end consumers and increase prices for farmers.**

DATA AS A PUBLIC GOOD

CHROME IAS

- **The decline in marginal cost of data clubbed with an increase in marginal benefit means that the optimal quantum of data that society should consume is much higher than before.**
- **Data carries some of the characteristics of public goods. It is non-rivalrous, i.e., consumption by one individual does not reduce the quantum available for others.**
- **Private sector investment in data-related endeavours is higher than ever before. In fact, in the last two decades, the world has witnessed the emergence of companies, such as Facebook, Amazon, Instagram, etc., who earn revenue exclusively from people's data.**
- **But, there are several areas where data is not as ubiquitously harnessed and used. Consider, for example, the agriculture market.**
- **While the private sector has done an impressive job of harnessing some kinds of data – the kind that can be converted into a private profit – government intervention is required in other areas where private investment in data remains inadequate.**
- **In the endeavour to create data as a public good, it is very important to consider the privacy implications and inherent fairness of data being used.**

DATA IN POLICY MAKING

CHROME IAS

- **Combining of disparate datasets can be extremely useful in obtaining the necessary richness required to design and implement welfare policies.**
- **Data needs to cover a critical mass of individuals/firms so that comparisons and correlations can be assessed among individuals/firms to generate useful policy insights.**
- **Data must have a long enough time-series so that dynamic effects can be studied and employed for policymaking.**

OPEN GOVERNMENT DATA

data.gov.in
Open Government Data (OGD) Platform India

- **The Union government's Open Government Data platform allows citizens to access a range of government data in machine-readable form in one place.**
- **The portal allows union ministries and departments to publish datasets, documents, services, tools and applications collected by them for public use.**
- **The platform also includes citizen engagement tools like feedback forms, data visualisations, Application Programming Interface (APIs) etc.**

- **Open data not only helps government officials make better decisions but also gets people involved in solving problems.**

DATA FOR THE WELFARE OF CITIZENS

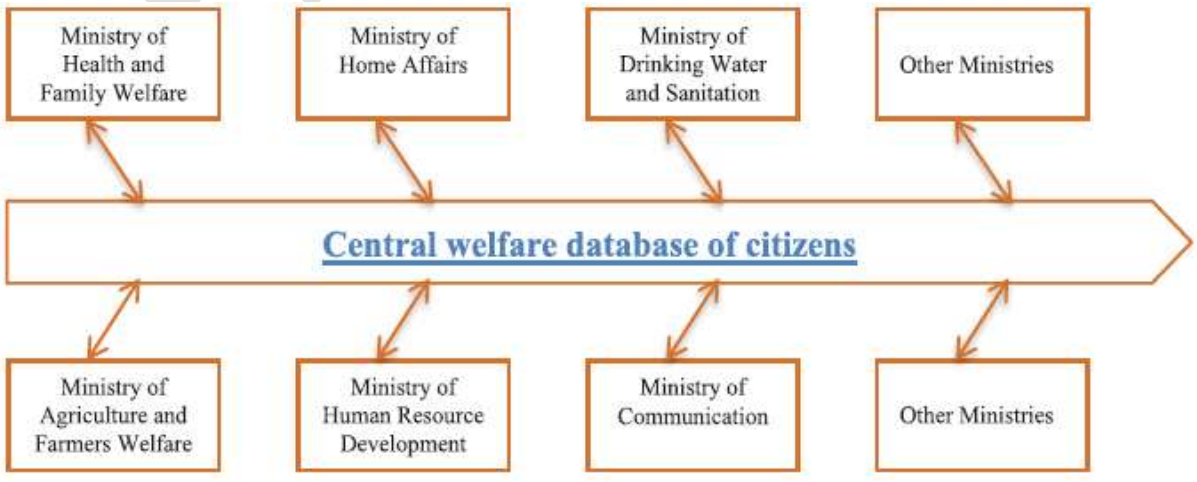
- **Administrative data: Governments hold administrative data for mainly non-statistical purposes. Administrative datasets include birth and death records, crime reports, land and property registrations, vehicle registrations, movement of people across national borders, tax records etc.**
- **Survey data: Survey data, on the other hand, is data gathered predominantly for statistical purposes through systematic, periodic surveys. For example, the National Sample Survey Office conducts large-scale sample surveys across India on indicators of employment, education, nutrition, literacy etc.**
- **Institutional data: Institutional data refers to data held by public institutions about people. For example, a government run district hospital maintains medical records of all its patients.**
- **Transactions data: Transactions data are data on an individual's transactions such as those executed on the United Payment Interface (UPI) or BHIM Aadhaar Pay. This is a nascent category of data but is likely to grow as more people transition to cashless payment services.**

CHROME IAS

SAMAGRA VEDIKA INITIATIVE

- **The Telangana Government's Samagra Vedika initiative gives a flavour of the potential benefits of integrating data sets.**
- **The initiative links around twenty-five existing government datasets using a common identifier – the name and address of an individual.**
- **Seven categories of information about each individual were linked in this aggregation exercise – crimes, assets, utilities, subsidies, education, taxes and identity information.**
- **Each individual was then further linked to relatives such as spouse, siblings, parents and other known associates.**
- **The initiative also puts in place all the necessary safeguards to preclude any tampering of data or violation of privacy.**
- **The right to add or edit data in the database varies by ministry or department. A given department can only write data for select fields – the motor vehicles department cannot, for instance, manipulate data relating to education, even though it can view the data.**

DATA ACCESS FIDUCIARY ARCHITECTURE



- **Each department of the government is responsible for making available the data they hold as a data provider.**
- **These departments must take care to appropriately treat private data and public data with the standards they require.**
- **This data is then made available through a Data Access Fiduciary to the Data Requestors.**
- **Data Requestors may be public or private institutions but can only access the data if they have appropriate user consent.**
- **The Data Access Fiduciary themselves have no visibility on the data due to end-to-end encryption.**
- **Such a model puts user consent in the centre of the government's initiative to make Data a Public Good.**

TRANSFORMING INDIA'S DATA INFRASTRUCTURE

Gathering data	•Digitize existing paper-based data •Initiate digital data collection at source
Storing data	•Initiate real-time storage for select data •Reduce time lag between collection and data entry
Processing data	•Build capacities of govt. bodies to analyse data •Involve private sector in analysis and insight generation
Disseminating data	•Create scheme dashboards •Open district-level dashboard to the public •Open data from third party studies to the public

DIGITIZE INDIA

- **Government departments upload scanned copies of paper records on the Digitize India platform.**
- **These scanned documents are shredded into snippets with meaningful data.**
- **These snippets are randomly served to digital contributors. Digital contributors are citizens who volunteer their time on the portal.**
- **Upon receiving snippets, the contributor reads the information and types it into a data entry portal.**
- **All converted data are verified against the corresponding snippet. Correct entries earn their contributors reward points, which can either be redeemed for cash or donated to the Digital India initiative.**
- **Once all snippets corresponding to a particular document are converted into digital data, the platform reassembles the document in digital form and supplies it back to the government department**
- **Any Indian citizen with an Aadhaar number can participate as a digital contributor.**
- **Citizens are incentivised through reward points and recognition as 'digital contributors' and may even earn certificates as 'Data Entry Operators'.**

CHROME IAS

- **The program also features a mobile app so that even citizens without a laptop or desktop computer can participate.**

NATIONAL HEALTH REGISTRY

- **For Swachh Bharat to transform into Swasth Bharat and eventually Sundar Bharat, citizens' health is paramount. Prevention is far more important in this endeavour than cure.**
- **A national health register, that maintains health records of citizens with all the necessary privacy safeguards can go a long way in enabling health analytics for predictive and prescriptive purposes.**
- **Such a national health register would be identified using a citizen's Aadhar.**
- **As a doctor can access the medical history of a patient from this national health register, this facility would be especially useful in emergency/trauma cases and can potentially save several lives.**
- **The various components of this register can include databases for**
    - **hospitals and public health centres,**
    - **surveillance of syndromes,**
    - **immunization information systems,**
    - **electronic laboratory reporting, and**
    - **sub-registries for key diseases requiring intervention such as diabetes, hypertension, cancer, AIDS, etc.**

CHROME IAS

- **Anonymized data from the register can be sold to private parties for analytics, which would then enhance prevention by offering predictive and prescriptive knowledge.**

WAY FORWARD

- **Through Aadhaar, India has been at the forefront of the data and technology revolution that is unfolding.**
- **As data for social welfare may not be generated by the private sector in optimal quantity, government needs to view data as a public good and make the necessary investments.**
- **The benefits of creating data as a public good can be generated within the legal framework of data privacy.**
- **Going forward, the data and information highway must be viewed as equally important infrastructure as the physical highways.**
- **Such a stance can help India leapfrog to utilise the benefits of technological advances for the welfare of its people.**

CONCLUSION

In the spirit of the Constitution of India, data "of the people, by the people, for the people" must therefore become the mantra for the government.

CHROME IAS

CHROME IAS