

Time: 20th June 2022, 10.00-16.00
Venue: Friedrich-Ebert-Foundation,

The Data Problem

The world today faces what can be described as a ‘data problem’, which is that the world’s data – its most important resource – is increasingly getting concentrated in fewer and fewer hands. This issue is perhaps the lead cause of many, if not most, of the ills of the digital economy that policy-makers are grappling with across the world. Within nation states, what this means is that by virtue of ‘owning’ and controlling the data-based intelligence of any given sector, one or two digital platforms unilaterally set the rules of interaction for all the economic and social actors in that sector, such as Uber vis-a-vis transport, Amazon for e-commerce, Facebook for digital social interactions, and so on. This data concentration is leading to unsustainable and unfair economic outcomes for small enterprises and workers, in addition to causing various social and cultural problems. At the geopolitical level, this ‘data problem’ manifests in the form of concentration of data power (including platform and AI power) at the two poles of the US and China, which power is instrumental to govern not just economies but societies as well. As a result, all other countries, including the European ones, face exclusions and dependencies that are painfully reminiscent of colonial times.

Data Infrastructures

The only viable solution to the ‘data problem’ is to make the non-rivalrous resource of data available in an equitable manner to all those who legitimately need it. Most of this data gets produced as a by-product of accessing various digital services, whereby a ‘tragedy of commons’ leading to its under-production is highly unlikely. Sharing this key resource would maximize economic output as well as promote fairness and equity. After all, most of society’s data is about and of the people and the communities that contribute such data, and they should have primary rights over it.

The term 'data infrastructure' has been turning up frequently in various policy documents of the EU and India. 'Data infrastructures' are active systems that make data available to a large number of actors on a real time basis, dynamically observing a number of conditions and constraints, as well as special access provisions. They are a kind of common facility underpinning and making possible a wide swathe of economic, social and other public activities. It is akin to how industrial age infrastructures like roads, electricity, banking, etc. made various industrial activities possible, and prevented their being concentrated in the hands of a few. Rendering them as infrastructures meant that they were made available in a non-discriminatory manner to all, and their exclusive private use was legally banned. Placing that 'infrastructure' framework over data shines light over the path to solving the seemingly wicked 'data problem'.

EU's ambitious GAIA-X program is about providing a distributed and federated data infrastructure for different sectors (with 'X' denoting the particular sector). On the other hand, India's take on data infrastructures is its 'public digital goods' initiative called the India Stack, whereby decentralised and interoperable protocols are used to provide a set of foundational digital services, which includes data management. It is now being built-over and implemented in different sectors. Some very interesting similarities and differences exist between the approaches taken by the EU and India, respectively.

Data infrastructures need data rights and data laws

First and foremost, there is a need to enact enabling or procedural laws like the EU's Digital Governance Act. Such enabling laws allow actors to voluntarily share their data in a safe manner while maintaining control over their data (the EU calls such control over one's data as 'data sovereignty', a term frequently found in GAIA-X related documents, for instance).

However, such voluntary sharing still does not address the original problem of a few global platform corporations collecting and hoarding most of society's data, and using it exclusively. Given the infinite competitive advantage such data provides, they are not willing to share this data even with the best of enablements, not even for a price. In such a scenario, how to get this extensive and crucial data held by a few corporations into public data infrastructures, without which the latter may not be of much use? One possible solution is to use the doctrine

of 'essential facilities' from competition law. However, this approach, while important, has significant limitations in the digital area.

There is now an increasing recognition of a need to give people, groups and communities legal rights to reclaim their data in another party's possession, as well as to control, use and move it. Since locking-in their data with a single service provider can potentially lead to an exploitative situation, data infrastructures provide an alternative where people's data is handled in a secure manner in their best interests and also made available to a larger set of competing service providers. Consequently, it is hoped that people would prefer such transparently-managed data infrastructures and employ their data rights to move their data into such facilities. The Digital Governance Act is largely about ensuring such conditions towards development of data infrastructures.

The EU's draft Digital Markets Act lays out some rights for the business users of gatekeeper platforms vis-a-vis their data collected by these platforms. The draft Data Act, on the other hand, institutes some general economic rights over data for those who help generate data. However, the latter is mostly restricted to data from IoT, or what can be called as non-human non-personal data (NPD) (as against anonymised personal data which is human NPD). (The underlying assumption, that individual holders of data rights – whether relating to personal or non-personal data – would invest into not only controlling their data but also making it available to different service providers, may need to be interrogated in the context of real-life digital situations where the rights holder is faced with irresistible offers from digital corporations of highly useful services, most of which are either free or subsidized. In this regard, the experience with portability rights granted under the EU's General Data Protection Regulation is not encouraging. The hope that an MSME owner – as a more rational and relatively empowered economic agent – will behave in any different manner with regard to IoT data generated in its workplace may also not entirely hold.)

One of the most valuable data is aggregated anonymised data (or human NPD) that helps predict human behavior at scale, and thus influence and control it. Leaving out this most valuable data from the framing of economic rights (as done by the Data Act) means that monopoly global platforms continue to have an exclusive and unconstrained use of it. This would not allow the fulfillment of the declared aim of EU's data strategy, and the legislative

proposals arising from it, which is to enable EU's digital industry become a global leader. Interestingly, the draft report of India's Committee of Experts on Non Personal Data Framework (NPD Report) primarily focuses on human NPD (though it also deals with non-human NPD). Since the subject of human NPD can by definition only be a collective, the report proposes collectives rights to such NPD of the relevant group or community. Such collective rights are then employed as the basis for developing data infrastructures, containing both human and non-human NPD. However, collective rights may be much easier to institute than to implement. The development of the necessary institutions in this regard may be a long and perilous path.

Different but complementing approaches of the EU and India

Leveraging its industrial strength, the EU's hopes to cash in on what it calls the 'second wave of digitalisation' (the first relating to data about people, and the second relating to data about things, machines, etc). Such data is largely yet not with platforms, and the EU's approach hopes that by providing rights to the relevant data generators through its draft Data Act, they can be nudged to direct their data to common data infrastructures or spaces. GAIA-X develops the implementing systems for this purpose.

Having huge digital skills and entrepreneurial energy, India, on the other hand, is focused on enabling its digital start-ups in different sectors. The main problem of these start-ups is that most of the key sectoral data, as well as the means for its continual collection, lies with incumbent dominant platforms. And access to such data is often key to any new digital venture. And India's NPD Report's answer to this is data trusts, through which large data collectors are mandated to share such infrastructural data upon meeting certain conditions.

It is not that India does not need to protect its manufacturing industry and MSMEs in the fast changing digital context, or the EU can do without addressing the data needs of its start-ups that need both human and non-human NPD. Although starting at different ends, as more and more aspects of the 'data problem' are sought to be addressed, EU and Indian approaches are likely to converge towards a common ground.

India's NPD report partly addresses the needs of small enterprises to control and manage their data (for instance, of farmers through data cooperatives). EU, on the other hand, is

recognizing the need for mandated sharing of anonymised personal data, as seen in its recent proposed Health Data Spaces Regulation. Similar regulation will evidently be needed for ‘populating’ other sectoral data spaces with all the required data – from sectors like transport, agriculture, and education to finance, tourism, and more.

The EU employs market failure and competition enforcement as the main legal basis for its data sharing mandates, while India’s NPD report relies on collective rights and community resource governance frameworks (taking from governance frameworks for community resources like traditional knowledge and natural resources). Both approaches have their advantages as well as significant limitations.

A more coherent and comprehensive legal basis for data sharing mandates may need to be thought out, which is most appropriate for a digital economy/society and the role of data in it.

Time to think holistic – Exploring a digital industrial policy framework

The ‘data problem’, as the main obstacle to a robust, fair and sustainable digital economy and society, is being approached from many perspectives in a piecemeal manner. This is understandable given we are still in the early stages of digital transformation. But it is also important to recognize that digital society structures that entrench into strong dominance are especially difficult, if not impossible, to undo. As has been aptly said, in a digital society, ‘*code is law, and architecture is policy*’. A deformed digital society architecture – beset with the ‘data problem’ – is well upon us, creeping further every day that the default digital/data arrangement is allowed to reign. It is time to begin looking at all the pieces and approaches together in what can form a coherent and effective digital industrial policy, with society-wide data availability and accessibility as its centerpiece. This is as much an imperative for the EU as for India, as it is also for the rest of the world.

The workshop proposes to examine and discuss the different approaches of the EU and India with regard to data infrastructures and data rights and laws. The objective is to help policymakers make well-considered and deliberate choices in their respective contexts to best deal with the opportunities and challenges posed by the digital economy/society, many of which are centered around the ‘data problem’.