Responsible Data Sharing for Public Good

A Review of Theoretical Bases and Policy Tools

by
Siddharth Manohar

THE DATA ECONOMY LAB

The Data Economy Lab is a partnership between Aapti Institute and Omidyar Network India

Acknowledgements

The author is grateful for the important feedback on this paper from Tripti Jain from Internet Democracy Project and Smitha Krishna Prasad from the Centre for Communication Governance. The paper is made possible by the broader guidance on the subject from Omidyar Network, the Data Governance Network with IDFC Institute, and colleagues at the Aapti Institute.

Contents

1 Introduction

The Evolution of Data Sharing

Across the world, data today exists in silos—scattered across sectors. Technology companies collect data on user behaviour online; governments collate data on schemes, programs, policies; and third party or data brokers, to whom data trickles over time, then further commodify it. Data—for example, location data collected from mobile phones—is used by technology companies to conduct analytics and offer more personalized services, by the government to offer more data-driven policy support and by third parties who sell this data to other businesses. At the back of all this are ad hoc data sharing agreements, which make it possible for this data to move from one entity to another.

There is growing realization that data accrues value when shared, combined and analyzed—generating more nuanced insights into the behaviour, choices and concerns of individuals and communities. Contemporary data sharing projects focus largely on municipal collection, storage and usage of data, such as for smart cities and in relation to public transport agencies: the X-tee data sharing system of Estonia, and Transport for London are good examples.[1] Data sharing across sectors may be leveraged for the wider benefit of user groups who are also the source of the data. The Humanitarian Data Exchange, for instance, is an open platform for sharing data across countries during crises and with organizations such as the UN Office for the Coordination of Humanitarian Affairs (UNOCHA).[2]

However, but for small and scattered efforts, data sharing, both in form and potential, remains largely under-explored. The aforementioned ad hoc patterns of data sharing have resulted in some cooperation between stakeholders, primarily through open data, but are now reaching a flashpoint in global discourse. There is also a realization that open data is insufficient, and there are multiple data sets and types that need to be and can be shared but cannot be thrown open due to privacy concerns. The varied and context-specific patterns of data sharing merit a study of the terms on which this sharing is being carried out. Who initiates the sharing? What are the trade-offs in sharing data? Do some use cases demand data sharing more than others? What are the models of sharing? What are the incentives guiding it?

1 Data Exchange Layer X-tee, Estonian Information System Authority, accessible at https://www.ria.ee/en/state-information-system/x-tee.html.

2 Humanitarian Data Exchange, UNOCHA: https://data.humdata.org/.

Given the somewhat overwhelming increase in the number of global data sharing models, there is need for an appraisal of sustainable and just versions of this regime, anchored in data rights. However, the objectives, rules, and governance of data sharing remain haphazard across contexts as well. A synthesized approach to governing data sharing, its practices and models, can help clarify and chisel the applicability of data stewardship as a method to unlock the value of data while protecting the rights of owner individuals and communities.

Data Stewardship and its Role in Evolving Frameworks for Data Sharing

The concept of data stewardship is taking root in research, across jurisdictions, and can provide answers to some of the questions. A governance layer based on principles of managing data in the interest of users, and realigning of incentives in relation to the usage of big data, constitutes a core tenet. It emerges from the immense social value that attaches to data. Traditional models of data storage and sharing are witnessing a shift towards data sharing within and amongst stakeholders, in the form of open data, data exchanges, and similar tools. The specific approach varies significantly according to sector, application, nature of data source, control exercised by participants, and other factors influencing data governance.[3]

Data stewardship seeks to evolve a model of sharing data that enables accountability and user interests,[4] but does not itself depend on a specific theoretical framework of data or a method of data comprehension in order to govern it sufficiently. It can be applicable to sovereign data, or data as a pooled resource, or data as a private resource. The current analysis seeds a discussion on the theoretical grounding of data stewardship to establish a common language and set of assumptions guiding regulatory and policy options to make data stewardship a viable regulatory choice.

This paper studies the evolving landscape and the basis of the new patterns of data sharing—the underlying thought processes, the instruments of implementation, and the objectives. The issue of prevention of data misuse—due to function creep, unauthorized sharing, surveillance measures and flouting of regulations through information asymmetry—is more relevant today than ever.[5] The paper studies the reasons for, and modes of, data sharing, and aims to synthesize an initial set of guiding principles for data stewardship. The legal and governance instruments that operationalize data sharing throw into relief the limits and methods of organizing data sharing and management. Hence, also included is a study

3 "Data Collaboratives: Leveraging Private Data for Public Good", Stefaan G. Verhulst, Andrew Young, Michelle Winowatan, and Andrew J. Zahuranec, GovLab, 2019, accessible at http://www.thegovlab.org/static/files/publications/data-collab-report_Oct2019.pdf.

4 "Understanding Data Stewardship: Taxonomy and Use Cases", Siddharth Manohar, Astha Kapoor, and Aditi Ramesh, Aapti Institute, 2020, accessible at https://www.aapti.in/blog/stewarding-non-personal-data/.

5 "Why Privacy is an Antitrust Issue", Dina Srinivasan, New York Times, 2019, accessible at https://www.nytimes.com/2019/05/28/opinion/privacy-antitrust-facebook.html.

of *(i) different philosophical bases for data sharing; and (ii) laws, policies and standards that support them.* The analysis and summation of the policies seeks to capture the thrust of their intent and supporting legislation.

This study of data sharing governance extracts the principles governing these structures and practices, i.e. data stewardship. This in turn creates a comparison between the rational basis for data sharing and the principles that in documented regulation actually guide the practice. The paper uses this comparison to proffer revised principles for data sharing which accurately reflect the underlying rationale for sharing data across stakeholders in the first place.

II

Structure of the Research Paper

The paper first engages with theoretical frameworks to manage data. It looks at perspectives on data—as a resource, as a matter of right, as an asset. These overlapping and divergent views are already prevalent, applied and understood in different contexts. Here, they are arraigned for comparison.

Then, each theoretical framework and its applications in the real world are explored. Next are the implications of competing theories for models and governance frameworks that correspond to them, and the legal and governance landscape of data distribution and control. An attempt is made to better frame the following questions in the light of existing work on the subject:

- *What is the theoretical basis for undertaking the practice of data sharing?*
- *What is the thrust of the intent behind these policies and legislation?*
- *What is their probable impact on data sharing patterns and their effects, given the current state of play and the restrictions it imposes?*
- *Can the developing approach impact data in order to promote social good?*
- *What principles can form the grounding for sustainable data sharing for beneficial purposes?*
- *How can these principles be applied to data stewardship?*

This will throw up a reading on whether and how the employment of these frameworks helps in achieving the stated objectives. Where directly replaceable by one another, they can be compared in achievement of similar objectives; and when achieving different outcomes, they may be able to overlap and be utilized together as well.

III

Theoretical Frameworks Guiding Data Sharing

Ideas that seek to reorganize data sharing patterns are the focus of this landscape portrayal and analysis of data sharing models. Emphasis is laid on models that seek to change the basis or structure of data sharing, and renegotiate current market practices.

Data Commons

In the contemporary design of the digital market, digital services harvest data as a byproduct and develop it over time to create a central intelligence system whose utility increases with greater input of data.[6] Technology companies are thus facilitated in understanding consumers better and providing tailored services; but they are also empowered to influence behaviour—as through targeted advertisements. The nature of this system makes data collection exclusive to each company carrying out such a service, like customer data on e-commerce websites or online behaviour data on social media platforms such as Facebook. This pattern incentivizes centralization of intelligence through its technical and economic design. Data and intelligence here are an economic resource concentrated in a single point of control. This accumulation of data harbours a competitive advantage and is safeguarded by business confidentiality and licensing contract practices.[7]

An alternative approach is that of the commons, a public resource freely accessible for the community. In Elinor Ostrom's seminal work, *Governing the Commons*, the term 'common-pool resource' is said to refer to "a natural or man-made resource system that is sufficiently large as to make it costly (but not impossible) to exclude potential beneficiaries from obtaining benefits from its use".[8] The premise of a data commons is that data is sourced from and is relevant to specific communities. Data on people living in a particular region, for instance, would be most relevant to those individuals, and for people and issues connected to them; data generated by a company through its services would be relevant to the company as well as to its customers from whom the data has been sourced. Data in this framework is seen as part of a "digital intelligence system" that has tangible social and economic impact through its role in organizing the distribution of and access to resources. Fundamentally, this approach recognizes the value of data beyond its use for the individual, and focuses on the collective—the data of an individual is only valuable when combined and analyzed along with the data of millions of others.

6 "Data and Digital Intelligence Commons", Parminder Jeet Singh, Data Governance Network, 2020, accessible at http://datagovernance.org/files/research/ITFC_Parminder_Data_Commons_-_Paper_2.pdf.

7 "Competition and Data Protection Policies in the Era of Big Data: Privacy Guarantees as Policy Tools", Nicola Jentzsch, accessible at https://fpf.org/wp-content/uploads/2016/11/Jentzsch_Ident_Workshop_Paper_2016_V8_FINAL-I.pdf.

8 Elinor Ostrom, *Governing the Commons: Evolution of Institutions for Collective Action*, 1990.

Given that intelligence is derived from collectives, the data should be used in the interest of these collectives. The interest of the community is represented by the commons—the structure of which needs to be determined in the digital realm. This requires not only policy, but legal articulation of the community's claim over resources based on information it has generated. This common pool resource of an intelligence system may also require sharing of data on sustainable licensing terms.

The commons is already used in management of resources in which, like aggregated data, a lot of users simultaneously have a stake in the management and usage.[9] Limited implementation of the role of this format in data regulation is seen in the draft of India's Personal Data Protection Bill, 2019, which introduces the idea of "Consent Managers",[10] analogous to the monitors of the commons in Ostrom's framework. Consent managers are agents for users to delegate the power to consent to sharing data with third parties. This framing, while avoiding any novel basis for sharing data, inserts an intermediary in the process to mediate between collectors of user information and third parties who wish to access that information, and operates based on consent provided by users.[11]

The European Union (EU) has enacted one of the most significant measures towards the idea of a data commons, through the Open Data Directive.[12] The legislation focuses on data collected by public sector entities and data collected through public funding. The legal basis for this reorganization of data remains fairly straightforward. The data available to state departments and official bodies is mandated to be made freely available for re-use. Wherever data is not protected by proprietary or privacy barriers, it is required to be made available for public use. This legal principle relates back to the idea of the data commons where an existing resource is stewarded for the public good.

The European Commission (EC) has identified high-value data sets to be made available amongst member states, free and with accessible API. The sectors of mobility, meteorology, statistics, corporate governance and geospatial data are covered.

Data Sovereignty

The concept of data sovereignty relies in turn on the concept of sovereignty, which involves supreme control by nation states over a territory, independent from other sovereigns.[13] In terms of data flow,

9 "The Future of the Commons - Beyond Market Failure and Government Regulation", Elinor Ostrom, Christina Chang, Mark Pennington, and VladTarko, Institute of Economic Affairs, 2012, accessible at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2267381.

10 "The Personal Data Protection Bill, 2019", Indian Ministry of Electronics and Information Technology (2019), retrieved April 24, 2020 from https://www.prsindia.org/sites/default/files/bill_files/Personal Data Protection Bill, 2019.pdf.

11 "Building safe consumer data infrastructure in India: Account Aggregators in the financial sector", Malavika Raghavan and Anubhutie Singh, Dvara Research, 2020, accessible at https://www.dvara.com/blog/2020/01/07/building-safe-consumer-data-infrastructure-in-india-account-aggregators-in-the-financial-sector-part-2/.

12 "European legislation on open data and the re-use of public sector information", European Commission, 2020, accessible at https://ec.europa.eu/digital-single-market/en/european-legislation-reuse-public-sector-information.

13 "Litigating Data Sovereignty", Andrew Keane Woods, Yale Law Journal, 2018, accessible at https://www.yalelawjournal.org/pdf/Woods_i233nhrp.pdf.

states can control the physical networks within their territories and their operation, as well as the intermediaries governing the traffic. Sovereign control also extends to the people and companies operating within the territory of the state. Therefore, the state may compel individual actors, including companies, to comply with laws and other requirements tailored to sovereign interests. The issue of the extent to which companies in possession of data may resist or cooperate with requests is dealt with in the paper excerpted below:

"...Data is just another globally distributed good, and as such its treatment by sovereigns and among sovereigns should abide by the usual rules of foreign affairs and international law. In the final analysis, if we choose indifference to deference, and allow ideals of internet cosmopolitanism to cloud our thinking, then states will eventually assert their sovereign differences anyway, and through worse means."[14]

Towards the end of this excerpt, Woods argues that the internet and digital resources should in fact be amenable to working with state requests—for the reason that a lack of cooperation may lead to more coercive measures by states, with worse outcomes for individual actors as well as the open internet as a whole. Measures might include data localization, involving restriction of all data generated in a country to servers located within that country—ensuring state control over all speech and communication online.[15]

It is important to recognize data sovereignty as part of a larger framework of sovereign interests in constant operation and interplay with other such forces of political power, both from other sovereign states as well as companies that may exercise power and influence over decision-making which may stray into the political realm, subjects normally considered the sole purview of nation states and their institutions. A push for data sovereignty represents the political operation of states that choose to inform policy on data through the expression of sovereign interest, as opposed to private interests of companies collecting data. Expression of data sovereignty therefore necessarily takes a different form—while companies focus on protecting IPR and related private rights that secure control over their resources, data sovereignty is represented by a top-down approach by nation states, both on an international policy platform and through domestic law, as seen in following examples.

This principle is being instituted in a number of ways across the globe. Russia, for example,[16] is exerting control through measures such as

14 "Litigating Data Sovereignty", Andrew Keane Woods, Yale Law Journal, 2018, accessible at https://www.yalelawjournal.org/pdf/Woods_i233nhrp.pdf.

15 "Data Localisation in China and Other APEC Jurisdictions", Scott Livingston and Graham Greenleaf, Privacy Laws & Business International Report, 2018, accessible at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2895610.

16 "Russia's push back against big tech has major consequences for Apple", Josh Nadeau, TechCrunch, 2020, accessible at https://techcrunch.com/2020/02/04/russias-push-back-against-big-tech-has-major-consequences-for-apple/.

compulsory-to-install applications.[17] The installation of applications on devices ensures a certain degree of data collection in parallel with digital platforms. While this measure directly targets devices and the data generated by them, there are other measures as well. The UK is instituting a digital platform tax[18] on big tech platforms. This remains a traditional financial tax, but targets the same set of companies that employ business models for strengthening their resources of proprietary data sets.[19] France has also enacted a tax on similar lines,[20] with the rule targeting tech companies within the country.[21]

Examples of implementation of data sovereignty abound but for the purposes of analyzing the mode of data sharing that they envisage it is useful to look at policies that (re)design data flows and platforms for data sharing, as in the case of the European Data Strategy Paper.[22] The EC released a policy strategy document in April month outlining plans for the creation of a data-agile economy. Building on the Open Data Directive, the policy seeks to work towards creating a cross-sectoral data governance framework that describes the processes and structures for data sharing across its member states. Elements of cross-border data flow, interoperability, and common standards are central to the exercise. The policy also focuses on creating an administrative structure to enable the process of collaboration and sharing data at a sectoral or cross-sector level. This structure will be governed by a non-interested entity trusted with pursuit of public good objectives, reflecting the function of a steward.

The policy also seeks to aid the creation of "data spaces",[23] in order to operationalize its aims. Data space is described as a common platform for data to be shared, based on a common governance framework and standards. The organization of data sharing spaces is intended to spur business innovation by making data available to SMEs, and create social benefits in areas such as environmental research.

In the example drawn from India, the Personal Data Protection Bill, 2019,[24] alongside an in-depth report (the Srikrishna Committee Report) charts the

17 "Russia: Regulators to crack down on US Big Tech in 2020", Competition Policy International, 2020, accessible at https://www.competitionpolicyinternational.com/russia-regulators-to-crack-down-on-us-big-tech-in-2020/.

18 "UK to impose digital sales tax despite risk of souring US trade talks", Alex Hern, The Guardian, 2020, accessible at https://www.theguardian.com/media/2020/mar/11/uk-to-impose-digital-sales-tax-despite-risk-of-souring-us-trade-talks/.

19 "UK finally takes on arrogant tech giants with digital services tax", Nils Pratley, The Guardian, 2018, accessible at https://www.theguardian.com/uk-news/2018/oct/29/uk-digital-services-tax-budget-facebook-google-amazon.

20 "Création d'une taxe sur les services numériques", French Parliament, 2019, accessible at http://www.senat.fr/espace_presse/actualites/201904/creation_dune_taxe_sur_les_services_numeriques.html.

21 France passes controversial tax on tech companies, Colin Lecher, The Verge, 2019, accessible at https://www.theverge.com/2019/7/11/20690253/france-digital-services-tax-google-facebook-tech-companies.

22 A European strategy for data", European Commission, 2020, accessible at https://ec.europa.eu/info/sites/info/files/communication-european-strategy-data-19feb2020_en.pdf.

23 "Stakeholders Dialogue on Common European Data Spaces", European Commission, 2019, accessible at https://ec.europa.eu/digital-single-market/en/news/report-european-commissions-workshops-common-european-data-spaces.

24 "The Personal Data Protection Bill, 2019", Indian Ministry of Electronics and Information Technology (2019), retrieved April 24, 2020 from https://www.prsindia.org/sites/default/files/bill_files/Personal Data Protection Bill, 2019.pdf.

vision for an operational data economy.[25] The Bill and the report adopt the framework of data fiduciaries and a consent-centric regulatory approach, and seek to set up a framework enabling companies to collect and process data in a predictable system based on user consent. However, one of the provisions of the draft Bill contains a mandatory clause where the central government, through consultation with the central Data Protection Authority, may order acquisition of any data that falls outside the legal definition of personal data.[26] The wide-ranging language leads to a blanket claim by the government to all data not containing identifiers of an individual. This also includes anonymized data, which may be modified by having such identifiers removed, though the nature of the anonymization process has not been further described.

The Srikrishna Committee Report also attempts to identify "community data" as a natural resource and defines it as "a body of data sourced from multiple individuals, over which a juristic entity may exercise rights".[27] Non-personal data and community data are gaining increasing relevance in the Indian discourse on data regulation. The earliest significant regulatory effort in this regard was the National Data Sharing and Accessibility Policy, 2012.[28] The policy set out guidelines for governmental agencies with respect to data handling and sharing. Intending to make public data more openly accessible, it prescribed processes for access to data collected and controlled by government agencies with varying levels of restrictions.

The concept of community data also finds mention in the 2019 draft for the National E-Commerce Policy released by the Department for Promotion of Industry and Internal Trade.[29] The policy attempts to make commercial data sets generated in India available to Indian companies and SMEs. It attempts to do this through measures such as provisions on data localization and mandatory data sharing.

On sharing of aggregated data, the Indian government has constituted a committee of experts under Kris Gopalakrishnan to 'deliberate on data governance framework'.[30] The Srikrishna Committee Report had signalled that specific regulations on non-personal data would be required, and the setting up of this Ministry of Electronics and Information Technology (MeitY) committee is an initial step in this direction. Further clarity on this issue is expected once future plans and activities of the committee are known. The

25 "A free and fair digital economy: Protecting privacy, empowering Indians", Committee of Experts under the Chairmanship of Justice B.N. Srikrishna (2018), page 45, accessible at https://meity.gov.in/writereaddata/files/Data_Protection_Committee_Report.pdf.

26 Section 91, "The Personal Data Protection Bill, 2019", Indian Ministry of Electronics and Information Technology (2019), retrieved April 24, 2020 from https://www.prsindia.org/sites/default/files/bill_files/Personal Data Protection Bill, 2019.pdf.

27 "A free and fair digital economy: Protecting privacy, empowering Indians", Committee of Experts under the Chairmanship of Justice B.N. Srikrishna (2018), accessible at https://meity.gov.in/writereaddata/files/Data_Protection_Committee_Report.pdf.

28 "National Data Sharing and Accessibility Policy", Indian Department of Science and Technology (2012), retrieved October 25, 2019 from https://nsdiindia.gov.in/nsdi/nsdiportal/meetings/NDSAP-30Jan2012.pdf.

29 "Draft National E-Commerce Policy", Department for Promotion of Industry and Internal Trade (2019), retrieved October 25, 2019 from https://dipp.gov.in/sites/default/files/DraftNational_e-commercePolicy_23February2019.pdf.

30 "Office memorandum: Constitution of a committee of experts to deliberate on data governance", Ministry of Electronics and Information Technology (2019), retrieved October 25, 2019 from https://www.medianama.com/wp-content/uploads/data-governance-framework.pdf.

trend across these measures is to maximize the legal claims of government over data sets controlled by companies and individuals within the territory of India.

Private Corporate Data Rights

THE TRADITIONAL MODEL for organizing data, this approach rewards data collection with exclusive rights to determine the terms of its usage and sharing. This understanding of data sharing is based on certain rights of parties and the exercise of these rights. The regime here is vaguely analogous, though not quite similar to that of intellectual property, where the usage of a person's property can be licensed to third parties on the owner's terms. In the case of data, companies share data that they have collected and processed with third parties for specific purposes and objectives.

Individual user rights take the form of claims against the company collecting and/or using the data. All data processing and sharing is done according to terms consented to by the user. However, these terms may not always be strictly enforceable against the processor of data.[31] They nonetheless serve as guiding practices, and users may claim damages in case of violations or otherwise undue harm as a result of data processing. Regulators have awarded penalties where the terms consented to have been significantly violated.[32]

Companies build data sets that become proprietary information—and thereby a part of their private assets, governed like other moveable assets of the firm. Usage of this asset is licensed to third parties through agreements between companies, forming the primary mode of data sharing in the market today.

As with intellectual property, rights to usage of data can be licensed to third parties by the company that generates the data, through agreements. Depending on the applicable laws in the sector and jurisdiction in question, and the type of data, this sharing of data with third parties is still subject to conditions of consent at the point of primary data collection and possible restrictions on purpose of usage.[33] For example, while laws may require disclosure of data for legal proceedings, these disclosures are communicated to users with a privacy policy or a document outlining the terms of usage.

It is relevant to note that this is demonstrably the most operational understanding of data sharing currently. Data is shared through bilateral

31 "The Clicks That Bind: Ways Users "Agree" to Online Terms of Service", Ed Bayley, Electronic Frontier Foundation, 2009, accessible at https://www.eff.org/wp/clicks-bind-ways-users-agree-online-terms-service.

32 "Facebook to pay $5bn fine as regulator settles Cambridge Analytica complaint", Rob Davies and Dominic Rushe, The Guardian, 2019, accessible at https://www.theguardian.com/technology/2019/jul/24/facebook-to-pay-5bn-fine-as-regulator-files-cambridge-analytica-complaint.

33 "Personal Data: The Emergence of a New Asset Class", World Economic Forum 2011, accessible at http://www3.weforum.org/docs/WEF_ITTC_PersonalDataNewAsset_Report_2011.pdf.

and multilateral contracts, with terms of access and usage rights of parties clearly laid out.[34] Proprietary data sets are built for larger and more systems of data processing, operating at scale for a larger variety of functions. Diversification of functions and increase in scale can be deployed by companies to maximize profits and grow in new markets across geographies and sectors.[35] To contextualize the larger drive towards expansive private aggregation of data and infrastructure built along this model, it is important to understand the incentives at play for companies and how this drives the development of intelligence systems. Factors of political economy mould this process; however, the arguments made in this paper are restricted to the impact of this paradigm on data sharing and its uses, without prescriptions for larger problems of incentives within what authors such as Zuboff would refer to as surveillance capitalism.[36]

34 "Ownership of Personal Data in the Internet of Things", Vaclav Janecek, Computer Law and Security Review, 2017, accessible at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3111047.

35 "Sources of Tech Platform Power", Lina Khan, Georgetown Law Technology Review, 2018, accessible at https://georgetownlawtechreview.org/sources-of-tech-platform-power/GLTR-07-2018/.

36 "Big Other: Surveillance Capitalism and the Prospects of an Information Civilization" Shoshana Zuboff, Berkman Center for Internet and Society, 2015, accessible at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2594754.

IV

Other Policy Tools Used For Data Sharing

Thus far we have appraised theoretical approaches to data sharing that organize principles of data sharing based on a particular conception of data— as a private asset, a national resource, or part of a commons. There are, however, some policy tools that work with the existing framework to reorganize the pattern of sharing data and information, without a fundamental divergence in how data is conceptualized.

Compulsory Licensing

Compulsory licensing in the intellectual property rights regime is a form of information sharing that has an existing legal basis and mandate.[37] The existing structure for sharing information under compulsory licensing is present under a number of provisions in patent law across jurisdictions. A compulsory license gives third parties the right to make use of the patented technology in addition to the patent holder, in exchange for compensation to the patent holder on fair, reasonable and non-discriminatory terms.[38]

This forms a structure of compulsory sharing of information (for instance, information on drug manufacturing) in direct pursuit of a public good objective. The basis for this sharing is the need established by a public need—such as the treatment for an epidemic.[39]

The principle of compulsory licensing is to be seen in the Agreement on Trade Related Aspects of Intellectual Property Rights (TRIPS)—under Article 31 of the document—presented as "other use without the authorization of the right holder", referring to the patent holder.[40] This legal tool is used in public health emergencies and in cases where there is evidence to show that the market is not benefiting from the holder's lack of usage of the patent.[41] It formed the basis of making production of the treatment for liver and kidney cancer in India open to health service providers other than the original patent holder.[42] Other examples abound across continents, in Brazil, Mozambique, and Germany.[43]

37 "Compulsory Licensing in India", Nayanikaa Shukla, Mondaq, 2019, accessible at https://www.mondaq.com/india/Intellectual-Property/772644/Compulsory-Licensing-In-India.

38 "Compulsory licensing, price controls, and access to patented foreign products", Eric Bond and Kamal Saggi, Journal of Development Economics, 2014, accessible at https://www.wipo.int/edocs/mdocs/mdocs/en/wipo_ip_econ_ge_4_12/wipo_ip_econ_ge_4_12_ref_saggi.pdf.

39 "TRIPS and pharmaceutical patents: fact sheet", World Trade Organization, 2006, accessible at https://www.wto.org/english/tratop_e/trips_e/factsheet_pharm00_e.htm.

40 "TRIPS and pharmaceutical patents: fact sheet", World Trade Organization, 2006, accessible at https://www.wto.org/english/tratop_e/trips_e/factsheet_pharm02_e.htm.

41 "Compulsory Licensing of Patents in India", Anubhav Pandey, iPleaders, 2017, accessible at https://blog.ipleaders.in/compulsory-licensing-patent/.

42 "India Grants First Compulsory Licence, For Bayer Cancer Drug", Intellectual Property Watch, 2012, accessible at https://www.ip-watch.org/2012/03/12/india-grants-first-compulsory-licence-for-bayer-cancer-drug/.

43 European Patent Office Report Compares Compulsory Licensing Practices By Country, Intellectual Property Watch, 2019, accessible at https://www.ip-watch.org/2019/03/01/european-patent-office-report-compares-compulsory-licensing-practices-country/.

A relevant qualification here is that this principle, as a codified legal provision, remains applicable only to intellectual property in the form of a patent and not to other intellectual assets, including data. The principled basis of sharing informational assets for public health emergencies can nonetheless prove useful in evaluating how data may be shared for humanitarian purposes.[44] The question of how better to aid humanitarian crises merits a further look at principles that can guide the use of technology tools for public good.[45]

Data Trusts in the UK

DATA TRUSTS IN THE UK have garnered significant research in the last few years to enable fiduciary responsibility to be applied to data stewards responsible for organizing data for sharing and usage towards specific purposes.[46]

A 2017 report commissioned by the UK government as part of its push towards development of artificial intelligence (AI) resulted in a host of recommendations on how to enhance the sector. They included the setting up of data trusts—defined in the document as "not a legal entity or institution, but rather a set of relationships underpinned by a repeatable framework, compliant with parties' obligations, to share data in a fair, safe and equitable way".[47] While the purpose of this policy was in part to enable data sharing initiatives for existing national institutions, there was also an element of enabling data-holding third parties to share data in a "fair, safe and equitable way".[48]

Further research on data trusts since then has explored the legal and governance implications, stating that the existing legal framework of trusts is unsuited to governing data sharing as it requires a clearer legal framework around data as property—which does not as yet exist. Rather, it encourages the formation of structures where the rights and interests in data can be strengthened through fiduciary duties of a body that stewards data.[49]

The concept of a trusted intermediary for data sharing has spawned research on the forms and roles of such a body in varying cases of stewarding data, with differing purpose, size, and legal structure.[50]

---

44 "The State of Open Humanitarian Data: What data is available and missing across humanitarian crises", Center for Humanitarian Data, 2020, accessible at https://reliefweb.int/sites/reliefweb.int/files/resources/StateofData2020.pdf

45 "From Principle to Practice: Humanitarian Innovation and Experimentation", Sean Martin McDonald, Kristin Bergtora Sandvik, & Katja Lindskov Jacobsen, Stanford Social Innovation Review, accessible at https://ssir.org/articles/entry/humanitarian_innovation_and_experimentation.

46 "Algorithms in Decision Making", House of Commons Science and Technology Committee, 2018, accessible at https://publications.parliament.uk/pa/cm201719/cmselect/cmsctech/351/351.pdf.

47 "Growing the Artificial Intelligence Industry in the UK", Dame Wendy Hall and Jérôme Pesenti, 2017, accessible at https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/652097/Growing_the_artificial_intelligence_industry_in_the_UK.pdf.

48 "Algorithms in Decision Making", House of Commons Science and Technology Committee, 2018, accessible at https://publications.parliament.uk/pa/cm201719/cmselect/cmsctech/351/351.pdf.

49 "Data Trusts: Legal and governance considerations", Open Data Institute, 2019, accessible at https://theodi.org/wp-content/uploads/2019/04/General-legal-report-on-data-trust.pdf.

50 "Designing decision-making processes for data trusts: lessons from three pilots", Mark Bunting &Suzannah Lansdell, Involve UK, 2019, accessible at http://theodi.org/wp-content/uploads/2019/04/General-decision-making-report-Apr-19.pdf.

V

Private Database Rights Versus Data Sovereignty Versus Data Commons: Evaluating Competing Interests

Let us examine how ideologically divergent frameworks on data rights interact, specifically with regard to their approaches to data sharing.

Private data rights form the currently dominant design of data sharing. Rights to share are inhered in the compiled databases controlled or accessed by various companies. The firms who control data license specific forms of access to third parties. Data commons, on the other hand, locate data rights in the group of people from whom the data is obtained. Any secondary use of this data is based on negotiation of these rights, or the use is beneficial to the group in some manner—and this forms the basis for data sharing. Data sovereignty puts the nation state at the centre of data claims. These claims may operate within as well as across national borders, as in cases of governments directing global platforms to modify content or provide information on specific users for violation of domestic laws.[51]

The Tension between Companies and States: Private and Rights versus Sovereignty

Data sovereignty, for all practical purposes, operates in conjunction with private database rights: that is, state claims on data sourced from digital platforms often take the shape of claims made against existing private databases, as opposed to primary data collection. Primary data collection by states may serve specific municipal and regulatory purposes, but these remain supplementary to offline governance measures. The state claim on big data, however, seeks to influence the contours of the digital market, and derive economic benefits for its constituents. This is evidenced by a number of policies across jurisdictions, such as the Digital Single Market policy in the EU, supplemented by its B2G Data Sharing Expert Group, whose aim is to create a framework of data sharing by companies across the continent along with pooled spaces where this data can be accessed by smaller companies and municipal authorities. Similarly, India's drafts of the e-commerce policy and the Personal Data Protection Bill contain clear provisions that indicate the intent to create legal bases for widespread data sharing across companies.

The trend of data localization battles illustrates[52] the relationship that these two approaches share: data sovereignty and private database rights are engaged in continuous negotiation wherein private databases continue to amass information while under the purview of domestic laws across jurisdictions, and the conflict comes into visibility when a nation state

51 "Litigating Data Sovereignty", Andrew Keane Woods, Yale Law Journal, 2018, accessible at https://www.yalelawjournal.org/pdf/Woods_i233nhrp.pdf.

52 "The Localisation Gambit: Unpacking Policy Measures for Sovereign Control of Data in India", Arindrajit Basu, Elonnai Hickok, and Aditya Singh Chawla, The Center for Internet and Society, 2019, accessible at https://cis-india.org/internet-governance/resources/the-localisation-gambit.pdf.

makes a claim on information from these databases. The claims that lead to these disputes may be narrow in terms of a specific piece of information, or much broader, like an in-principle claim to information, as in the case of data localization drives across the world.[53]

Institutional Rights versus Collective Claims: The Clash of the Commons

THE DATA COMMONS CONCEPT too comes into conflict with these ideas. At first glance, while it may be seemingly more aligned with data sovereignty, this is demonstrably not the case. For, although sovereign power may act as a tool to enforce a claim for a data commons, these ideas remain opposed in principle and primary objective. The idea of data sovereignty lends primacy to state claims on data—in other words, claims on behalf of the government of a given territory. The data commons entity, on the other hand, makes a claim on the group that forms the source of the data—the extent of the claim differs, and the claim itself also may directly contest against a claim by the state.

In Ostrom's work, the commons is described as a framework where there is no single centralized authority with primary claim over resources. Instead, the participants and beneficiaries of the common pool of resources nominate a monitoring agency acting as a neutral entity to mediate between the parties sharing the common pool resource.[54] An important distinction here is that no claim of a party is made even through the monitoring agency; the terms of sharing are agreed directly between the participants of the sharing pool, and any claims are made directly to the parties. The monitoring agency acts as a keeper of records and information, only serving to mediate communication and cooperation.

Ostrom's framework contrasts with a system of centralized distribution of resources, where the authority to control and distribute resources flows from the power of the sovereign. This is in line with the idea of data sovereignty, and can be seen articulated in the examples of data localization and the Indian conception of "community data" represented in the Srikrishna Committee Report.[55] These policies rely on the premise of the claim of the state, on behalf of the collective, negotiating against the claims of individuals as well as companies. The conflict between the data commons and data sovereignty therefore arises when there are competing claims between groups of people and the government that serves them.

In theory, this conflict may be resolved in cases where state policies accurately and effectively represent citizens' interests and institutional mechanisms are set up to allow negotiation in cases of competing claims. Ideally, the framing of individual and collective rights, enforceable against the state, should be sufficient for securing user interests in this conflict.

53 "Data Nationalism", Anupam Chander and Uyen P. Le, Emory Law Journal, 2017, accessible at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2577947.
54 "Governing the Commons: Evolution of Institutions for Collective Action", Elinor Ostrom, 1990, pg. 125.
55 Srikrishna Committee Report, 2018, pg. 45, accessible at https://meity.gov.in/writereaddata/files/Data_Protection_Committee_Report.pdf.

The claim of the state, however, may take the shape of eminent domain, and the overriding claim of the state has been argued in the Srikrishna Committee Report as well. A more rigorous study of this issue, though, requires an objective comparison of the competing interests of the citizens and the state, and to ask what mechanism, if any, can resolve this conflict of interests.

A clearer line of division, however, exists between data sovereignty and corporate data rights. The disputes that arise as a result of these conceptions of data have already been referred to, such as with data localization conflicts and governmental requests for specific information. Nonetheless, sovereign power could in theory still serve other ends within its own borders. The question is whether data whose access has been negotiated through sovereign power can be accessed on the terms of individuals or groups from whom it arises—putting the power of setting the terms of data usage back at the source of the data (also the end users)—and making possible the leveraging of data for purposes of negotiated public benefit, by decentralizing control over data usage. Using this double-pronged analysis of public good and user control, the ideal system of data stewarding would be able to leverage data from its various sources and forms, and use it for public good through terms negotiated with user groups from whom the data is sourced. This would require the employment of a non-interested party that would be able to balance these competing interests and manage data sharing and usage while reflecting and respecting the outcome of negotiation of these interests and their regulation.[56]

56 Wanted: Data Stewards: (Re-)Defining the roles and responsibilities of Data Stewards for an age of data collaboration, GovLab, 2020. http://www.thegovlab.org/static/files/publications/wanted-data-stewards.pdf.

VI Conclusions And Further Questions

This paper has looked at two modes of analyzing data sharing: the theoretical underpinnings of how data is conceptualized and the justifications for sharing data that flow from these theories, and the various legal instruments that reflect these understandings of data and lay down principles of data sharing between specified stakeholders. The analytical portion of the paper pits these ideas against the others, using their articulation in research as well as in policies, and seeks an approach that accounts for the limitations of the approaches while serving the common objective of enabling data sharing for public good and securing the interests of user groups from whom the data is sourced.

The conflicts between the theories of data sovereignty, data commons, and private data rights led to the understanding that while the first two seek a reorganization of data sharing and access, the terms on which this is carried out and the interests served are nonetheless divergent. The sovereign power may not always align with the interests of user groups, though sovereign power may still be held to standards of serving public good through accountability measures. At the same time, both ideas remain in conflict with private data rights of companies while data sovereignty remains an ongoing negotiation on multiple fronts through battles over access to information and control over data flows. If the negotiating principle of data sovereignty can be leveraged, in theory objectives of public good may be served through the creation of a commons, though the system that can enable this is a pending question. Private data rights and the data commons are more directly opposed ideas, but if there is a clear basis for a claim from user groups for control over data sourced from them, there is in theory a possibility for a data resource stewarded for the benefit of the user group.

If data can be leveraged on the basis of any of these ideas, it is important to organize it on the lines of accessible public benefit. This requires the employment of a non-interested party, a steward, who would manage data sharing and usage in a manner accommodating the competing interests and the terms of negotiation and regulation applicable to the sharing and usage of data. This steward would need to be structured as a non-interested party politically and economically, to be able to act as a neutral party[57] enforcing policies and mediating divergent interests of varied stakeholders.

57 Wanted: Data Stewards: (Re-)Defining the roles and responsibilities of Data Stewards for an age of data collaboration, GovLab, 2020. http://www.thegovlab.org/static/files/publications/wanted-data-stewards.pdf.