Jain
Family
Institute

GLOBAL
PULSE

Data Commodification

Opportunities and Challenges
Findings from the JFI/UN Global Pulse January 2022 Peer Exchange

Arden Ali & Friederike Schüür

Contents

Jain
Family
Institute

GLOBAL
PULSE

Executive Summary

We are transitioning away from an industrial economy oriented toward manufacturing to an informational economy oriented toward the production, accumulation, and processing of information. As a part of this transition, personal and non-personal data is becoming a factor of production alongside land, labor, and capital: data plays an increasingly larger role in the development of products and services that generate surplus value. Brought forth by this transition, data commodification is an emerging phenomenon, a cause and consequence of data becoming a factor of production. The term 'data commodification' refers to the process through which data is turned into a good intended for exchange. This report details the promise and perils of data commodification. It identifies key questions and areas of work needed to reap the benefits of data commodification while reducing risks and potential harms.

The promise of data commodification is to disrupt current trends in data such as: (a) **monopolization**: data is controlled by just a few organizations who are poised to derive the majority of the benefits of the emerging informational economy; (b) **exclusivity**: marginalized groups are not well-represented in data, so products and services trained on data are less likely to serve their needs, and these groups lack the influence to direct the use of data to their benefit; (c) **exploitation**: the labor needed to fuel the informational economy (data collecting, cleaning, labeling, etc.) is undervalued and undercompensated and; (d) **underutilization**: while precise estimates vary, there is broad consensus that much data is never used.

By foregrounding the exchange value of data, and thus facilitating its use, data commodification promises to address the underutilization of data. Furthermore, data commodification may provide opportunities for small data holders (including individuals) to benefit from the commercial value of their data. Since commodification involves treating goods as appraisable property, it can empower small data holders who gain a new entitlement, insight into its value, and possible financial benefit. As such, data commodification promises to provide a check on monopoly power and a mechanism to make the informational economy more inclusive. Similarly, data commodification makes it possible to assess and appraise value-add in terms of data cleaning and labeling, and thus offers a path towards more appropriate compensation for data labor.

There is great potential to use data for social and humanitarian purposes. But (e) increasing **commercialization** means that the exchange and use of data is primarily oriented toward realizing its commercial value at the risk of excluding valuable applications in the interest of the public good. Data commodification is likely to exacerbate this trend. The use of data for social or humanitarian purposes today relies on data philanthropy, public-private partnerships, and open data initiatives. To date, these approaches have failed to truly unlock the benefits of data for social and humanitarian purposes. Data commodification presents a crucial need, and opportunity, to safeguard data as digital public goods by placing them into the digital commons.

Jain
Family
Institute

GLOBAL
PULSE

2

At the same time, there is risk of "corporate capture" of data commodification (i.e., agenda setting by corporate interests). Data commodification may entrench current monopolies (and exacerbate exclusivity and exploitation) if ownership models for data are not scrutinized. Furthermore, applying current approaches of individual ownership to data may be a bad fit for relational data such as genetic data, transactional data, and other data that are about more than a single individual. Future ownership models for data must accommodate the relational nature of data with consequences for how data can be commodified. In addition, there are concerns about privacy: differential economic pressures on small data holders (including individuals) may turn privacy into a luxury good (i.e. something only the rich can afford).

More foundationally, there is precedent for banning, or putting limits on, the commodification of certain things. Across societies, we have placed strict limits on the commodification of human bodies and body parts (surrogacy, organ donation, etc.). We have yet to determine whether there are types of data that are impermissible to commodify because they constitute excessive "self-commodification" and thus threaten human dignity, autonomy, and flourishing. Given the very personal nature of some data, these questions will remain pressing as data commodification proceeds.

This report is a follow-up to a peer exchange on *Data Commodification* organized by the Jain Family Institute and co-hosted by United Nations Global Pulse in January 2022. It is our hope that it will provide a foundation for applied research on data commodification and support a new community of practice dedicated to a vibrant, global, just, and inclusive informational economy that promotes the use of data for the broad benefit of humanity and the planet.

From the Executive Director of UN Global Pulse

Data defines our age. "The global economy has become a perpetual motion machine of data," according to a *Foreign Affairs* article in the Mai/June 2021 issue by Matthew J. Slaughter and David H. McCormick with the adroit title *Data is Power.* "Measured by bandwidth, cross-border data flows grew roughly 112 times over from 2008 to 2020." Data has commercial value, it is playing an increasingly larger role in the development of goods and services. It truly has become an input to production alongside land, labor, and capital as part of the still ongoing transition from an industrial to an informational economy.

The sustainable development goals (SDGs) recognize the pivotal role of labor and other key inputs for economic, social, and humanitarian development such as energy and set the global agenda for safeguarding access to these key resources. While the SDGs highlight the important role of data in tracking the implementation success, they are largely silent on the emerging role of data as an input to production. Combined with the current trend towards data monopolization—a select few holding exclusive digital keys to vast amounts of data—access to data (or lack thereof) is likely to emerge as a key driver of widening inequalities. We need new guidelines and goals for the responsible stewardship of data, including the distribution of commercial benefits, in order to avoid destabilizing inequalities, growing poverty, and conflict as called for by the UN Secretary-General in his 2021 report *Our Common Agenda.*

In addition to commercial value, data has social and humanitarian utility. The Covid-19 pandemic showcased the value of broad access data. Open access to genome sequences of the virus accelerated the development of the Covid-19 vaccine. Epidemic modeling informed by open access population statistics provides insight into the potential effectiveness of public health intervention to inform policy making. Data literally saves lives.

UN Global Pulse is the UN Secretary General's cross-pillar and system-wide hub for digital innovation, experimentation, and anticipation in support of UN priorities and for advancement of the UN Charter. We aspire to realize the social and humanitarian utility of data in collaboration with partners from across the UN family and beyond. At the same time, we strive to identify and mitigate risks and potential harms that may arise from digital innovation.

This report examines the current but largely overlooked trends towards data commodification, the process of turning data into goods intended for exchange. It charters a new and productive path towards identifying areas of concerns in our emerging data economy and paths towards truly realizing the promise of data as a digital public goods for our collective, sustainable futures.

I urge you to join this important conversation!

Robert Kirkpatrick, UN Global Pulse

From the Chief Research Officer at JFI

It's by now a truism that data is of central importance to today's world. Data fuels consumer applications ranging from thermostats to self-driving cars; agricultural planning and "smart farm" operations; industrial process control and supply chain logistics; pharmaceutical breakthroughs and medical treatments. Company valuations are increasingly driven by recognizing data holdings as valuable — sometimes the most valuable — assets. Across all business sectors, global spending on data and analytics is over $200 billion per year and growing fast.

Meanwhile, concerns over privacy, inequality, and abuse have become a mainstay of popular press coverage of data. Scholars and activists[1] have built a strong case that a data-driven-everything approach, without strong safeguards (i.e. restrictions on what and whose data, what's driven and who's driving it, what subject matter, for whose use, for whose benefit, and for what purpose) will, reliably and predictably, reinforce structural injustice and oppression, serve the interests of the powerful, and do harm to the most vulnerable.

So data is important. But what does that claim mean, exactly? "Importance" isn't so much a property as an ellipsis: something is important when it is causally influential for, inferentially informative about, or constitutively determinative of *something* (to be filled in by context). So when we talk about data's importance, we should be clear: importance for what? In just two paragraphs, we've contemplated luxury consumer satisfaction, human wellbeing, productive capacity, resource and infrastructure use, business spending, inequality, social justice, economic and political power, and more.

Our understanding of data as not just a technical object, but as a social, political, and economic one, is still in early stages. We are, in William Ogburn's terms, in a state of cultural lag, with developments in technology far outpacing those in both our collective understanding and our normative intuitions. Combined with the multiplicity of data's importance, this conceptual nascency creates a rhetorical trap with attendant intellectual risk: it is quite easy to (as in the first paragraph) point out (perfectly correctly) a set of ways (viz., monetary) in which data matters, treat them as the central ways in which it matters, adopt its indicators (e.g market share or effect on GDP) as natural metrics of importance, and cast other forms of importance, including data's implications for social injustice, as external side-effects. Perhaps more nefariously, such framing can introduce presuppositions that frame future discussion — presuppositions that may be difficult to recognize or challenge. In particular, market-centered framings of data tacitly introduce an assumption that data is a commodity, to be owned and traded, and naturalize this assumption for future conversations — all without explicitly considering the question of whether data is, or should be, a commodity.

This report takes on this critical challenge, naming this assumption and subjecting it to scrutiny. What forms does data commodification take, and with what effects? What social possibilities does it expose and foreclose? What guardrails can we build to ensure that commodification preserves the social value of data, and where should we consider limiting it? Foregrounding these questions and more, this report lays the foundations for a vital, timely, and exciting new line of applied research.

Jerome Hodges, Jain Family Institute

---

[1]Including Joy Buolamwini, Timnit Gebru, Safiya Noble, Cathy O'Neil, Virginia Eubanks, Issa Kohler-Hausmann, Andrew Selbst, Solon Barocas, Debbie Hellman, Clinton Castro, Katie Creel, and many more.

Jain Family Institute

GLOBAL PULSE

Data as an Emerging Factor of Production

Personal and non-personal data is playing an increasingly larger role in the development of products and services that generate surplus value. It is thus becoming a factor of production alongside land, labor, and capital and a vital component for economic development.[2] At the same time, data has the potential to improve the well-being of people and the planet. Data has social and humanitarian utility in addition to commercial value.

Land and labor were commodified as societies transitioned from an agrarian to an industrial economy through large-scale enclosures and propertization (and privatization) of land and the constructions of factories with assembly lines. Today, we are seeing the emergence of an informational economy oriented towards the production, accumulation, and processing of information.[3] With this transition, we are witnessing the commodification of data alongside the emergence of data exchanges (marketplaces for data) and a trend toward the propertization of data if not in legal terms then in practice.[4] This is reflected in data becoming an input to production.

Commodities are goods intended principally for exchange and commodification is the process that positions goods as commodities.[5] The candidacy of a good as a commodity is socially and culturally determined and moral considerations put limits on what may be subject to commodification. We have yet to answer these questions for data (see **Box 1: Commodification and the data economy**).

This report examines data commodification and asks how data should be governed as a new and emerging factor of production, and what lessons from past large-scale transitions may apply. This report is the follow up to a Peer Exchange on *Data Commodification* held in January 2022 organized by the Jain Family Institute and co-hosted by UN Global Pulse.

---

[2] In *Building the New Economy: Data as Capital* by Alex Pentland, Alexander Lipton, and Thomas Hardjono published by MIT Press in 2021, the authors argue that "data is now a full-fledged means of production, and consequently, we need to think about it as a new type of capital, along with human and financial capital".

[3] To clarify, the industrial and informational economy do not unfold in strict succession: they overlap and coexist for potentially long periods of time. The terms are also not meant to suggest that *immaterial* aspects are irrelevant in the industrial economy (brand recognition, for example, is a driver of value also in the industrial economy). Furthermore, manufacturing, the primary mode of generation of surplus value in the industrial economy, is increasingly digitized: data is playing a larger role in manufacturing, too. The primary point we wish to make is that with the transition towards the informational economy, data is playing an increasingly larger role in the generation of surplus value.

[4] In *Between Truth and Power: The Legal Constructions of Informational Capitalism* by Julie E. Cohen published by Oxford University Press in 2019, the author highlights the ongoing transition toward the informational economy and lays out how legal instruments (copyright, intellectual property, etc.) are reinterpreted to serve powerful corporate interests with regards to exclusive access to and effective ownership of data. We also adopted her definition of the informational economy.

[5] Not only goods but services can be, and are, commodified. See *Introduction: Commodities and the Politics of Value* by Arjun Appadurai published in *The Social Life of Things: Commodities in Cultural Perspective* by Cambridge University Press in 1986. For simplicity, we here use 'goods' to refer to both.

Jain
Family
Institute

GLOBAL
PULSE

> ### Box 1: Commodification and the data economy
>
> **Commodities** are goods (or services) intended principally for exchange. **Commodification** is the process that turns a good into a commodity. Goods are turned into commodities in certain situations, situations in which their use-value is highlighted for the purpose of exchange. It is thus more appropriate to speak about phases during which a good is treated as a commodity (the **commodity phase**).
>
> Goods enter a commodity phase when they enter a **commodity exchange**, a type of exchange that is object-centered, relatively impersonal, and highlights the use-value of goods (unlike a gift exchange). It is a normative question whether a good can and should enter a commodity phase, and if so, under what circumstances. It is not morally permissible to treat human beings as a commodity while it is permissible to treat their labor as such but frequently subject to restrictions (e.g., human and labor rights). It is generally frowned upon to foreground the use-value of art with the exception of art auctions, a commodity exchange where art enters the commodity phase under very specific circumstances that are generally accepted.
>
> Commodification usually leads to **property rights** being defined over the goods that are eligible to enter the commodity phase. In capitalist societies, commodification tends to lead to the **pricing** (financial appraisal) of these goods as well.
>
> **Data commodification** is an emerging phenomenon. The **data economy** is a global, digital ecosystem in which data is gathered, organized, and exchanged. Data exchanges provide opportunities for data to enter a commodity phase. The data economy and data exchanges are both the consequence of data commodification *and* the driver. As data commodification is an emerging phenomenon, we have yet to define the conventions for data to enter the commodity phase, whether or not it is permissible, and under what circumstances.
>
> —
>
> Our definitions are informed by *Introduction: Commodities and the Politics of Value* by Arjun Appadurai published in *The Social Life of Things: Commodities in Cultural Perspective* by Cambridge University Press in 1986.

While conceptual in nature, this report is applied in orientation. The Sustainable Development Goals (SDGs) recognize the pivotal role of labor (Goal 8: Decent Work and Economic Growth) as well as other key inputs for economic, social, and humanitarian development (Goal 7: Affordable and Clear Energy) for the well-being of people and the planet.[6] The SDGs highlight the important role of data for tracking the success of the implementation of the SDGs by key development variables (such as gender equality, etc.). However, the SDGs omit the governance of data as an

---

[6] For further information about the SDGs, please see here: https://www.globalgoals.org/goals/.

Jain Family Institute

GLOBAL PULSE

input to production. Similarly, *Our Common Agenda*, a 2021 report of the UN Secretary-General, details commitments to accelerate the attainment of the SDGs.[7] While the report mentions the need to protect data, it offers limited guidance on how to unlock the value and utility of data in the informational economy as a factor of production.

But, there is a need to ensure the responsible stewardship of data as an input to production to avoid potentially destabilizing inequalities, growing poverty, and conflict given the growing role of data for development. This includes the fair distribution of the commercial value of data and sustainable pathways to realize the social and humanitarian utility of data. This report examines data commodification as part of our ongoing transition towards the informational economy with an eye towards safeguarding our collective, sustainable futures.

Trends in Data and Consensus on a More Desirable Future
We note five trends in data today and examine them through the lens of data commodification.

First, the vast amount of data is currently managed and effectively owned by just a few organizations headquartered in just a handful of countries mostly in the Global North. The data economy threatens to become **monopolistic** with large data holders reaping outsized benefits. Data from sensors that track the movements and purchases of customers at Amazon Go stores has been used to optimize inventory management and the layout of Amazon Go stores. Data from sensors in Tesla cars is used to generate insight into driving behavior and optimize assisted driving features of Tesla cars. This monopolization affects both specific parts of the data economy and the economy as a whole. Since Google has nearly complete dominance in the market for search engines, for example, it holds a monopoly on data related to user search but it has also been able to monopolize more broadly within the data economy, using data collected from its email service to dominate the market for internet browsers. Organizations that manage vast data can leverage this data as an input to their own production of AI products and services. The surplus value they generate is almost exclusively to their and their stockholder's benefit. This contributes to and exacerbates existing inequalities within and across countries. Monopolies in the data economy are concerning. Given asymmetries in power, individual stakeholders are unable to effectively advocate for their own interests and protect themselves against violations of privacy, manipulation, and equal protection. Further, data monopolization has the potential to slow innovation.

Second, the benefits of data are liable to become more **exclusive**. This is partly a result of monopolization. But, even aside from monopolization, marginalized groups are especially vulnerable in the emerging informational economy. They are not well-represented in data. Any products and services trained on existing data are less likely to serve their needs well.[8] They

---

[7] *Our Common Agenda* by the UN Secretary General published in 2021 can be accessed here: https://www.un.org/en/content/common-agenda-report/.
[8] For example, in *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification* by Joy published in 2018 in the *Proceedings of Machine Learning Research*, the authors demonstrate that facial recognition works less well for

Jain Family Institute

GLOBAL PULSE

furthermore lack the influence to direct the use of data to their benefit, limiting their opportunity to directly benefit from data (i.e., not just through products and services). They are at risk of being excluded from the benefits of data use.

Third, data collection and use can be **exploitative**. Data use is exploitative if there are no mechanisms for returning value to data originators. The labor needed to fuel the data economy, like data collecting, cleaning, and labeling, is currently undervalued and undercompensated. Data labeling is a form of knowledge extraction used for future automation. Those who engage in data labeling are at increased risk of having their skills devalued due to automation-related limitations on employment and development opportunities.[9]

Fourth, more than eighty percent of all data collected is never used. Data will continue to be **underutilized** if current trends persist.[10] More importantly, the broad use of data for social and humanitarian purposes is an often missed opportunity, and we collectively pay the opportunity cost.

Fifth, current trends encourage the use of data for commercial purposes to the exclusion of other valuable applications of data. **Commercialization** is in tension with the idea of data as a digital public good. A contributing factor as to why the use of data for social or humanitarian purposes may be missed as data commodification proceeds.

While data commodification may be fueling some of these trends, it also represents an opportunity to disrupt or course correct them.

At the peer exchange, there was consensus on the need to disrupt these trends with a particular emphasis on the importance of sharing the benefits of the data economy more broadly. There was recognition that individual stakeholders currently lack the rights, tools, and bargaining power to control the use of data about them and are unable to ensure fair compensation for their data. There was agreement that individuals should be empowered to participate in the informational economy on their own terms, including their need for privacy. There is insufficient attention paid to barriers to participation in the informational economy, especially barriers facing those in low- and middle-income countries, which can stand in the way of benefitting from the goods and services developed using data. There are urgent questions about how to increase capacity for participation in the informational economy as well as the representation of the most marginalized

---

traditionally disadvantaged populations. These systems were the most accurate for white men and the least accurate for black women. Consequently, black women are likely to experience fewer benefits from these systems and instead may experience exclusion including denial of service.

[9] For an excellent discussion of this point, we refer readers to *Ghost Work: How to Stop Silicon Valley from Building a New Global Underclass* by Mary L. Gray and Siddarth Suri published in 2019 by Harper Business.

[10] The exact numbers for the % of data never used varies across sources. In 2012, the *Digital Universe Study* by the *International Data Corporation* found that only 1% of data is analyzed. In 2018, Gartner reported that 97% of data is unused. These numbers need to be taken with a grain of salt, they are frequently cited by consulting firms looking to advertise their digital transformation services.

Jain Family Institute

GLOBAL PULSE

in datasets. Finally, there is the question of trust. Successful exchanges require trust among parties involved in the exchange.There are open questions about the role and importance of trust in the data economy and data exchanges and the role of potential institutions operating on behalf of or in the interest of people. We return to these issues throughout this report.

Fair Realization of the Commercial Value of Data

The current trends towards monopolization mean that the vast amount of data is currently managed and effectively owned by a few organizations, the majority of which are headquartered in the Global North. These organizations leverage data as an input to their own production of AI products and services. The surplus value they generate is almost exclusively to their stockholder's benefit. This exacerbates existing inequalities within and across countries.

Individuals currently have few options to maintain control over their data, and even fewer options to derive commercial benefits other than access to products and services without clear insight into the value of the goods exchanged.[11] In this exchange, individuals are likely shortchanged as they derive a one-term benefit whilst the holders of their data can use and reuse their data for different purposes over time (i.e. repeat benefits). Individuals by themselves will also struggle to realize the commercial value of their data directly as Machine Learning (ML), which is at the heart of the AI capabilities of products and services that leverage data during their development, requires large or diverse data which single individuals do not tend to produce and have. Therefore, they cannot commercially benefit from their data in the way large data holders can just at a smaller scale.

Organizations sometimes cite privacy as the reason they control and limit access to vast troves of data they hold. It is worth noting, however, that limiting access to data also serves their commercial interest.[12] This means that data, a non-rival resource, is effectively managed as a rival resource as organizations look to position themselves as the dominant player in order to realize the commercial benefits of data. It is worth pointing out that this also limits broad innovation in machine learning and artificial intelligence. Collectively, we pay the cost of missed opportunities. We will return to the issue of privacy in the section <u>Data Commodification and Data Protection</u>.

The commodification of data foregrounds data as a valuable good for exchange and as such it also foregrounds the (commercial) benefits of data exchange (for examples of initiatives that foreground the value of data exchange, see **Box 2: Illustrative data sharing initiatives**). The movement towards data commodification, as well as the tools being developed, enable easy, frictionless exchange of data with minimal transaction costs and thus remove the practical barriers of large-scale data exchange. They provide an opportunity and foundation for the empowerment of small data owners (including individuals) to commercially benefit from their data by selling their data or pooling their data with others to be managed in their best interests; propertization and pricing of data is presented as an approach to promote fairer realization of

---

[11] The information asymmetry of the current exchange of data-for-services is described in *Data as Labor: Valuing Individual Contributions to the Digital Economy* in *Radical Markets: Uprooting Capitalism and Democracy* for a Just Society by Eric A. Posner and E. Glen Weyl. They use this information asymmetry to argue for the benefits and role of market mechanisms to empower individuals in expressing their preferences regarding data collection and, more importantly, use.

[12] We refer the reader to *Industry Unbound: the Insight Story of Privacy, Data, and Corporate Power* by Ari E. Waldman published in 2021 by Cambridge University Press for an excellent discussion of how the discourse on privacy is leveraged by corporations to limit access to data in the name of privacy but to advance their own corporate interests.

Jain
Family
Institute

GLOBAL
PULSE

commercial benefits. Small data holders are empowered by conceptualizing data as a good for exchange, and providing an infrastructure for pricing it, and exchanging it. Crucially, to truly fulfill their promise, these data exchanges need to facilitate data exchange at a *vast scale* and with *minimal transaction* costs to reach individuals and unlock the long-tail of data (small data owned by the many).

These infrastructures are being developed today, Ocean Protocol and Pool are just two example efforts, and they also support the management of data pools by individuals as part of data collectives, collaboratives, and trusts.[13] These data pools increase the bargaining power of individuals as large and diverse data is more valuable. In many ways, increasing availability of these new large-scale, frictionless data exchanges is driving data commodification.[14]

Data commodification as such, as a way to address monopolization and exclusivity, can be seen as an antitrust measure designed to facilitate broad access to and commercial benefit from a currently locked resource. This is not surprising. Historically, the push towards commodification is often spearheaded by entrepreneurs looking to unlock untapped (or undertapped) and often monopolized resources.[15] However, without data pools, there may be concerns about the ability of individuals to participate in data exchanges at eye level with corporations and other organizations. Union-like structures such as data pools for data offer a solution.

Data commodification also increases transparency in the exchange of goods. Currently, individuals hand over their data without knowing the value of it. Putting a price on data, as some advocate for, carries the potential to increase transparency and facilitates informed exchange. The capability to price data is in development. Proposals leveraging Automatic Market Makers (AAMs) and other tools from Decentralized Finance (DeFi) to use market mechanisms for the pricing of data, a traditionally tricky undertaking due to the non-rival nature of data; as data is not used up by use, usage-based pricing needs incorporate all potential future uses of data.[16] AAMs from DeFi offer a new solution to this issue firmly positioning data as a new asset class. This setup also opens up opportunities for compensating for data labor such as data quality improvements and thereby encouraging broad effort towards better data, something that is badly needed given the poor state of much data. This may offer ways to address the concerning trend toward data exploitation.

---

[13] For information on these organizations, please refer to https://oceanprotocol.com/ and https://pooldata.io/. Pool in particular foregrounds the value of data pools in the form of data trusts, unions, cooperatives, collaboratives, etc.

[14] Opportunity for exchange drives commodification as it increases the chances of goods to enter their commodity phase, see Box 1: Commodification and the Data Economy.

[15] In *Introduction: Commodities and the Politics of Value* by Arjun Appadurai published in *The Social Life of Things: Commodities in Cultural Perspective* by Cambridge University Press in 1986, the author notes that entrepreneurs excluded from certain goods tend to advocate for their commodification as a way to increase access to these goods.

[16] For an introduction to the challenge of data pricing, please see *Let's Talk About Data Pricing* by Erwin Kuhn published in 2019 on the Ocean Protocol blog https://blog.oceanprotocol.com/lets-talk-about-data-pricing-part-i-bbc9cf781d9f (accessed on March 10 2022).

Jain Family Institute

GLOBAL PULSE

Data commodification is not without concerns. First, it is important to clarify (initial) ownership of data. Without clarifying ownership, we risk entrenching the current ownership model of data owned by a few global corporations. Putting it bluntly, the movement towards data commodification carries potential to exacerbate monopolization and the push towards propertization of data may be, at least in part, motivated by current corporate interests.[17]

Second, the extent to which the individual ownership model can apply to data is unclear. Human genetic sequencing data may come from a single individual, but it is about a group of individuals. A picture taken by an individual may have a group of people in it. Who should be in the position to consent to data collection and use? Who should own the data? The individual model to ownership may be insufficient when it comes to data. Relatedly, some data carries the potential to produce broad societal benefits. Open access to genome sequences of the Covid-19 virus accelerated the development of the vaccine. Data has social and humanitarian utility. What are sustainable mechanisms for data as digital public goods?[18] We will return to this in section Safeguarding the Social and Humanitarian Utility of Data.

Third, there is a need to determine if all data should be commodifiable. Societies traditionally put limits on the commodifiable of goods and services from outright bans to regulation. The commodification of human beings is banned. The commodification of human organs and body parts is strictly regulated. Should all data be commodifiable? Data can be very personal (e.g. data on brain activity) since it may provide deep insight into someone's personality and traits and it is likely that these insights could be exploited commercially. More generally, there may be a need for special protection of data regarding our personhood and agency in the digital and virtual realm beyond the current definition and protection of "personal data" to safeguard human dignity, autonomy, and flourishing.There is a need for special protection also to ensure that privacy, with the ability to sell data, does not become a luxury good only accessible to the select few.

Fourth, current proposals towards the commodification of data and especially proposals that advocate for the pricing of data as a grounding for exchange are rooted in a neoliberal governmentality that foregrounds the utility of markets for the efficient management of resources. These proposals are not ideologically neutral. We need to be mindful of the fact that we export ideology alongside solutions and assess fit, risks, and potential harms as we do so.

Finally, there are concerns about the specific tools that underpin large-scale, frictionless data exchanges. These largely blockchain-based proposals enable auditable data transactions at scale without a central authority. We return to this issue in Data Infrastructures and Protocols.

---

[17] And we have learned from the commodification of land that unclear initial ownership leads to protracted problems in case of land over centuries with some of the most vulnerable being shortchanged.
[18] For an introduction to the concept of digital public goods, we refer the reader to the Digital Public Goods Alliance https://digitalpublicgoods.net/.

Jain Family Institute

GLOBAL PULSE

Box 2: Illustrative data sharing initiatives

There are targeted efforts to increase data sharing, use, and reuse. These include:

- The UN Global Pulse led **Global Data Access Initiative** (GDAI), a collaboration with the World Food Programme and the UN Development Programme and other public and private sector partners. GDAI strives to realize the social and humanitarian utility of data and puts emphasis on the importance of providing *responsible*, human rights-protecting access to data as digital public goods including privately held data.

- The European initiative **GAIA-X** "for Europe by Europe" strives to create a federated and secure data infrastructure that enables companies and citizens to share data in such a way that they keep control over them. The architecture of Gaia-X is based on the principle of decentralization following a common standard, the Gaia-X standard.

- **Ocean Protocol** looks to provide tools for the Web3 (blockchain based) data economy to empower data holders, including small data holders and individuals, to retain control over their data and also to enable them to reap commercial benefits.

- India's **Data Empowerment and Protection Architecture** (DEPA) is predicated on the notion that individuals should have control over how their personal data is used and shared. It is designed with the belief that agency over data could empower Indians with opportunities to improve their own lives.

- **Pool Data** is developing a Web3 platform to facilitate creation of data unions to give individuals more control over their data including data use, monetization, and sharing.

- The **Data for Common Purpose Initiative** (DCPI) by the World Economic Forum (WEF) seeks to create a flexible data governance model that allows for the combining of data from personal, commercial, and government sources, while still respecting privacy rights, to empower a variety of stakeholders while removing unintended policy barriers.

Collectively, these initiatives and similar ones drive data commodification.

Jain
Family
Institute

GLOBAL
PULSE

Safeguarding the Social and Humanitarian Utility of Data

The current trend toward commercialization of data in service of the data economy may dominate and crowd out the use of data for social and humanitarian purposes. However, data has the potential to improve the well-being of people and the planet (for examples, see **Box 3: Social and humanitarian applications of data**) and it is essential to encourage data use and sharing for these purposes alongside the commercial uses of data.

Governments currently rely on public-private partnerships to unlock the utility of data for social causes. A typical public-private partnership involves collaboration between government and private businesses to meet a broadly beneficial social goal. These arrangements take different forms. In some cases, public and private partners jointly offer financial contributions to the project. In others, the collaboration chiefly involves the provision of technology, in-kind technical advice, or joint technology development.

While public-private partnerships are in many ways desirable, it is unclear how they can facilitate the use of data for social and humanitarian purposes on their own. Generally speaking, these relationships are not clearly defined and largely opaque to the public. This can create difficulties in terms of accountability and responsibility for projects that take this form. In addition, when data is sourced from a privately-held firm, it may be difficult to properly assess the quality of data, including its sourcing and collection, and it may impose constraints on data use and sharing that may limit the extent to which the social and humanitarian utility of data can be realized.

Further, public-private partnerships risk public dependency on private data sources. Public sector agencies who depend on private partners for data may be unable to provide key services without private partners' assistance/cooperation. Other kinds of dependency may follow. When infrastructure and expertise is provided through the private sector, the public sector may fail to develop its own expertise and tools, which can leave it unable to regulate the private sector down the road. This increases risk of corporate capture.

Data philanthropy is the other approach through which data is put to social and humanitarian use. Some companies voluntarily donate data assets with governments and non-profit organizations to help improve public services. There is no doubt that philanthropic data sharing is key to the mobilization of data to humanitarian causes.[19] For this reason, UN Global Pulse has advocated for a "global data philanthropy movement" over the last decade since its inception. But there are limitations to this strategy. For one, some insights from data are somewhat serendipitous; the opportunity afforded by the data is only visible once the data is in hand. If donation is the primary

---

[19] Data philanthropy and data altruism is at the heart of the European Union's Data Governance Act designed to increase data sharing through the creation of data intermediaries. Please refer to: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52020PC0767

Jain Family Institute

GLOBAL PULSE

model through which data is put to social or humanitarian purposes, there will be missed uses. Further, it is also notable that companies have market incentives to restrict the extent of their data philanthropy.

Some have championed 'open data' as a tool for increasing social and humanitarian uses of data. However, the open data movement has stagnated in recent years.[20] And it is unclear that merely 'open data' is the right framing for the issue at hand. In one notable case, open access to land records was intended to empower poor citizens to view information that had been previously restricted, but the newly opened data was primarily used by real estate developers.[21] When data is merely open, there is no guarantee that the information will be used in ways that reduce inequalities and promote social good.

Truly and appropriately capturing the social and humanitarian value of data will likely require more than open data, public-private partnerships, and data philanthropy. Yet data commodification appears poised to make it even more difficult to secure the humanitarian and social benefits of data. But this juncture may also present an opportunity. If trends toward commercialization and exclusivity within the emerging economy are disrupted through data commodification, this movement could be leveraged to encourage social use. Put differently, data commodification is an occasion for the reconceptualization of data's relationship with the public good. There are two ways that data can be oriented toward the public good.[22] Data *for* the public good refers to the voluntary use of privately-held data to promote social good in the public sphere. The use of public-private partnerships and data philanthropy presuppose that data is fundamentally private, but can be put to use for public causes. But data *as* a public good refers to a more formal recognition of data as a non-rivalrous good whose benefits should be broadly accessible to all as part of a digital commons. The difference between data for the public good and data as a public good is subtle but important. In the former, use of data for social or humanitarian purposes is a pro bono or charitable choice, but in the latter, use of data for social good is an entitlement of the public as a whole.

Treating data as a commodity is compatible with treating some kinds of data as public goods. The work to reformulate data as a public good has already begun under the Digital Public Goods Alliance[23], who have set standards for recognition as a digital public good. It will be critical that the window of opportunity opened up by data commodification for repositioning data *as* a public good is not missed.

Putting data to use for humanitarian and social causes faces its own challenges. There can be especially high standards for data management when the data concerns people who are vulnerable

---

[20] See the OpenData Barometer at opendatabarometer.org
[21] See Bhuvaneswari Raman's 2012 article, "The rhetoric of transparency and its reality: Transparent territories, opaque power, and empowerment" in *The Journal of Community Informatics* 8(2).
[22] For this distinction, see Linnet Taylor's 2016 article, "The ethics of big data as a public good: which public? Whose good?" in *Philosophical Transactions of the Royal Society* 37: 2083
[23] https://digitalpublicgoods.net/

Jain
Family
Institute

GLOBAL
PULSE

along some dimension or targeted for poor treatment. Crises can change the sensitivity of data in ways that complicate data protection. Non-personal data about traffic patterns may be typically low risk, but may become sensitive in the context of violent conflict.

Even if data is treated as a public good, there is no guarantee that the appropriate kind of data will be available for social or humanitarian use. In a highly commercialized market for data, businesses are naturally discouraged from collecting information from individuals with less economic or social standing. As a result, safeguarding the use of data for social good will require that data collection from vulnerable groups is funded or otherwise incentivized. Further, those who lack regular access to sufficient, private, and secure connectivity to the internet are often ignored by standard data collection practices. In that way, increasing reliance on data to direct humanitarian projects can exacerbate existing inequalities between the rich and poor, a key challenge to tackle.

Box 3: Social and humanitarian applications of data

Some uses of data for social and humanitarian purposes include:

- In response to the Rohingya refugee crisis, the **UN Center for Humanitarian Data** and the **International Organization for Migration** used satellite images, drone footage, and other data sources to identify camps that were at greatest risk of flooding and intervened before the start of the rainy season.

- To help battle the spread of the Zika virus, the **Oswaldo Cruz Foundation** partnered with **IBM** to predict the spread of mosquitoes that carry the virus using rainfall, temperature, and humidity data.

- In collaboration with the **Red Cross**, **DataKind**, and local fire departments, **Enigma Labs** used publicly available data to identify homes least likely to have smoke alarms in New Orleans and directed local coordinators to prioritize outreach efforts to those homes.

- To help enable rapid and accurate analysis of satellite imagery to inform disaster response efforts, **UN Global Pulse** in collaboration with the **UN's Satellite Center** (UNOSAT) developed PulseSatellite, a web-based tool allowing users to interact with machine learning models for a range of analysis tasks using satellite imagery data.

- In collaboration with the **World Health Organization**, **UN Global Pulse** built a global radio monitoring tool to track discussions of COVID-19 to support the infodemic response in Africa (e.g., to decrease vaccine hesitancy) leveraging transcribed talk radio data.

- To support the UN High **Commission of Refugees** (UNHCR), **UN Global Pulse** used ship Automatic Identification System (AIS) data to study rescues of migrants and refugees in the Central Mediterranean.

Jain Family Institute

GLOBAL PULSE

Data Commodification and Data Protection as Countermovements

The growing importance of data and the role of data as an input to production has incentivized increasingly broader and largely indiscriminate data collection. Very large data sets have been created. Alongside the still growing data stockpile, there is heightened concern about data collection and use and the rights of individuals with respect to their information. This has launched protective countermovements to safeguard individuals from potentially excessive and exploitative corporate interests. The emergence of these protective countermovements is not surprising. There were similar efforts in the late 19th and early 20th century in response to the transition from the agrarian to the industrial economy.[24] As we transition away from the industrial to the informational economy, the rise of new protective countermovements is, in fact, to be expected.[25]

These countermovements aim to ensure proper articulation and protection of the interests and rights of people. They are important and can deliver meaningful benefits to people. However, any movement can overshoot its objectives and in fact (unintentionally) harm those it seeks to protect. Movements can also be "captured" by corporate interests. In this case, the movement continues to serve their stakeholders in appearance but the proposed measures are ineffective, largely symbolic, and serve to assuage concerns instead of resulting in substantive change.

Data Protection and Privacy Protective Countermovement
Data commodification can be regarded as a countermovement to data monopolization, exclusivity, exploitation, and underutilization. There are other, relevant countermovements that provide context for data commodification; most prominently, the protective movement in response to indiscriminate data collection centered on data protection and privacy. Data commodification may be regarded as a reaction to the perceived failure of the data protection and privacy movement, and there are several distinct ways that the data protection and privacy movement may currently be failing to deliver on its objectives.

First, data protection may limit opportunities to realize the social and humanitarian utility of data and the potential role of data as a digital public good. Governments today struggle to access data (particularly privately-held data) that could improve public services from public transportation and tourism to crisis prevention and recovery.

---

[24] This dynamic is described in *The Great Transformation: The Political and Economic Origins of Our Time* by Karl Polanyi first published by Beacon Press in 1944.
[25] Julie E. Cohen makes this point in *Between Truth and Power: The Legal Constructions of Informational Capitalism* published by Oxford University Press in 2019.

Jain
Family
Institute

GLOBAL
PULSE

Second, data protection may limit opportunities for innovation by not making data more broadly accessible as well as opportunities for entrepreneurs to develop AI-enabled products and services. While *meaningful* data protection and privacy is important, data protection and privacy are not infrequently given as key motivators for limiting access to data by data holders. While legitimate in many cases, as noted earlier, data protection and privacy can be used in service of corporate interests to limit access motivated by the commercial value of data.[26] The protective countermovement centered on data protection and privacy struggles may struggle, at times, with a form of corporate capture.

Third, the protection of data privacy in the United States and European Union is largely approached from the notice-and-consent angle: individuals are informed about data collection and asked for their consent (the details vary by jurisdiction). This approach has been criticized as largely symbolic and ineffective. Individuals cannot be reasonably expected to handle many notice-and-consent decisions a day as a condition to access digital services. Dark design patterns, the default set to the most comprehensive data collection, steer people to consent to broad data collection without realizing the implications of their consent. The protective countermovement may be ineffective and, again, appears to have been captured.

Data Commodification as a Response

Data commodification can be seen as a reaction to the perceived ineffectiveness of the data protection and privacy countermovement and corporate capture. Data commodification has the potential to unlock access to data for public service improvements and as digital public goods given proper safeguards. The European Union's Data Governance Act, a piece of legislation that fuels data commodification, recognizes the pivotal role of data philanthropy for these purposes and aims to put in place solutions to meaningfully increase data philanthropy. The aspirations of the data commodification movement also include the enablement of broad innovation and entrepreneurship by making data more broadly accessible.

It is interesting to note how innovation is used in this discourse. Current data holders argue for restricting access to data to foster innovation: restricting access to data is positioned as a key mechanism to incentivize data collection for (in-house) innovation. Limiting access to data, however, means fewer people are able to innovate with the data, the argument put forward by the data commodification movement. This points toward the need for a closer examination of the innovation narrative. The need for innovation is not uncommonly used to argue against movements and protective countermovements alike and it can be used to argue against legislation (to advance corporate interests) when regulation, in fact, can also spur significant innovation.[27] The

---

[26] See *Industry Unbound: the Insight Story of Privacy, Data, and Corporate Power* by Ari E. Waldman published in 2021 by Cambridge University Press.

[27] See Julie E. Cohen in *Between Truth and Power: The Legal Constructions of Informational Capitalism* published by Oxford University Press in 2019.

Jain Family Institute

GLOBAL PULSE

innovation narrative should not be taken at face value. We regard it as an open question whether data commodification will help or hamper innovation.

Returning to data commodification as a reaction to the perceived failure of data protection and privacy, aspirations for data commodification include the empowerment of individuals to enjoy the commercial benefits of their data and to engage in data philanthropy. As such, data commodification may be at the heart of finding a better tradeoff between openness (data collection, access, use, reuse, etc.) and data protection and privacy by delegating that choice to individuals (not regulators or corporations). As noted earlier, data pools (unions, trusts, cooperatives, etc.) are expressly designed to empower individuals and increase their bargaining position in the face of corporate interests. Data trusts in particular provide a powerful model for collective bargaining in this context. A data trust aims to pool individuals sharing data in order to directly address power-asymmetries.[28] In general, a trust is formed when one person (a settlor) vests a personal resource with someone else (a trustee), who has a duty to manage that resource in the interest of someone other than the trustee (a beneficiary). In the case of data trusts, data subjects are typically both the settlors and beneficiaries of the trust. Trustees would have a fiduciary duty to manage a data subject's information in ways that benefit the data subject. By pooling data under the direction of a trustee, individual data subjects gain leverage in negotiating for fairer, safer, and more transparent uses of their information, including the protection of their privacy interests. In short, data commodification may offer choice in face of the current failure of data protection. It can enable small data holders to pick options that are a good fit for their data protection and privacy needs while keeping day-to-day decisions to a manageable minimum through data pools.

Substantive questions about data pools remain. For example, it remains unclear how to individuate or diversify data pools to context, geographies, and populations. A broadly-based pool presumably retains more power for data subjects, but narrowly-based unions are naturally more sensitive to the special needs of beneficiaries. It is also an open question whether data pools can address the relational nature of some kinds of data, like genetic information or transactional records.

With regards to privacy, the data commodification movement, of course, has been criticized. Incentivizing the sale of data, in particular personal data, may lead to people giving up privacy for commercial benefit with particular pressure exerted on individuals susceptible to economic pressure. For example, data that provide insight into how people travel might be valuable for advertising firms to target the placement of physical ads, but if sold can also erode privacy (and may even expose individuals to risks). Some respond to this criticism by noting that privacy is ineffective in today's world anyway. While data commodification may not offer mechanisms to protect privacy other than individual preferences and choice, at least it provides a way to compensate individuals for the costs of privacy loss if they choose to share their information.

---

[28] See Sylvie Delacroix and Neil Lawrence's article, "Bottom-up data Trusts: disturbing the 'one size fits all' approach to data governance" in *International Data Privacy* (October 1, 2019).

Jain Family Institute

GLOBAL PULSE

However, the ability to sell data prompts questions about the rights and entitlements of individuals with respect to their data that, as part of the data commodification movement, are usually framed in terms of ownership. Data commodification naturally moves us towards the propertization of data. But, should all data be understood as property? What rights and entitlements accompany data as property? While the move to propertize data can be seen as an attempt to empower individuals, it may fail to recognize the relational nature of data (genetic data, transactional data, etc.). Future ownership models for data have to be able to accommodate the relational nature of data, or the idea of collective ownership of data as part of a (new) digital commons, with downstream consequences for how data can be commodified.

Data Infrastructures and Protocols

The commodification of data requires digital infrastructure, specifically, infrastructure that facilitates data exchange. This ranges from data storage to identity management and from data exchange platforms to data exchange protocols. Some of that infrastructure is available and has matured (e.g., infrastructure to store data) and it is part of the standard offerings from cloud providers. Similarly, APIs (Application Programming Interfaces) can be (and are) leveraged for data exchange. To fully realize the promise of data commodification, however, we require new infrastructure to facilitate data exchange with minimal transaction cost and to flexibly handle consent to data use (vs. collection) at scale.[29] This new infrastructure is currently being developed. Of course, data commodification is the cause and consequence of emerging technical capabilities and infrastructure; the availability of tools encourages their use.

A set of emerging tools for large-scale data exchanges foregrounds the benefits of decentralization and leverages the blockchain protocol. Blockchain-based data exchanges promise to dramatically reduce the transaction costs of data exchange. The blockchain coupled with smart contracts, which are snippets of code that represent (legal) agreements and automatically execute when certain conditions are met, enable parties to exchange without the need for intermediaries (such as lawyers). This opens opportunities for individuals to directly express and realize their data use preferences, access commercial benefits, and contribute to the use of data for social and humanitarian purposes at scale (intermediaries such as lawyers tend to introduce bottlenecks limiting throughput and scale).

The blockchain-approach also records all data about exchanges on the shared and distributed ledger (which is at the heart of blockchain). This has the potential to increase transparency and auditability of data access and use. Blockchain also allows the use of AAMs (Automated Market Makers), which position data as a new investment opportunity. Individuals can gain by investing in (i.e. backing) certain data. They are incentivized to back more useful, higher quality data over less useful, lower quality data as their returns increase with the usage of the data. As such, blockchain-based proposals put forward a discovery mechanism for usable, high quality data and, as mentioned earlier, incentivize data quality improvements. These are real solutions to pressing problems. The discoverability of quality data is a hard challenge that currently available infrastructure solutions (data storage, APIs, etc.) struggle to solve (mere preference or usability ratings are not honest signals and are thus less valuable).

One of the benefits of decentralization is that there is no central authority that can control or tamper with the exchanges enabled by the blockchain. This can be good in the case of potentially malicious central authorities. Control, however, can be an asset in case of malicious transactions enabled by blockchain. Control is also useful for curtailing transactions for other legitimate reasons.

---

[29] To enable small data holders to reap the benefits of repeat data use, consent needs to be given.

Effective data exchanges also require agreement on standards for formatting data, including the layout and interpretation of columns in structured data. Lack of standardization increases transaction costs and puts limits on the ease with which data can be used, thereby limiting the effective reuse of data. Incentivizing data access and exchange by leveraging market mechanisms encourages adherence to a shared standard, since data that is standardized is more useful and thus more valuable. There is real value in leveraging markets to address this issue. However, the market does not guarantee that the optimal solution emerges. It provides no guarantee that what emerges will respect the interest of data owners, data users, and those impacted by data and derived products. For example, it is commonplace to encode gender as binary in data. This practice ignores the interests of nonbinary groups by effectively erasing them from the data. This point generalizes. Markets generally risk favoring the interests of the dominant majority at the expense of many, including the most vulnerable.

In addition to infrastructure, tools, and standardization, it is important to recognize that the process of data commodification does not happen in a vacuum. It is socially, culturally, and historically embedded in ways that raise ethical and political questions. As noted above, data commodification is likely to reflect and exacerbate social and economic inequalities. From the perspective of commercial interests, data about those who lack social and economic power is likely to be considered less valuable. Individuals subject to economic pressure may also be willing to sell at a lower price. Furthermore, the incentives to carry out data collection for these groups may be fewer. Finally, certain forms of knowledge are more likely to be systematically ignored and undervalued, especially those forms that are difficult to standardize among more dominant forms of knowledge. The social embeddedness of data commodification poses challenges that blockchain-based solutions cannot address but may further entrench.

In summary, the considerable benefits of data commodification for individuals will not extend broadly unless careful attention is paid to barriers to participation in the data economy. Individuals who lack access to reliable and safe connectivity will certainly be blocked from enjoying any benefits from the new economy. Those who know little about distributed ledger technology or smart contracts may be discouraged from participation in a new opportunity that could improve their economic conditions markedly. And without established data trusts or unions, individuals have little chance of getting a fair deal from the exchange of information. Insofar as the motivation for data commodification rests on the benefits that will accrue to individuals, it is incumbent on those leading the charge to provide the social and economic conditions needed for those benefits to be shared fairly.

Conclusion and Moving Forward

As we transition toward the informational economy, data is becoming an input to production and increasingly subject to commodification. Data commodification may drive but also promises to disrupt current and concerning trends in data including **monopolization**, **exclusivity**, **exploitation**, and **underutilization**. However, the realization of the promise of data commodification is not guaranteed: movements can (and have) failed to meet their own objectives. Furthermore, data commodification is likely to exacerbate the trend towards **commercialization** and may therefore displace applications of data for the public good. In addition, close examination of data commodification and targeted action is needed to effectively realize the social and humanitarian utility of data while identifying, tracking, and mitigating against risks and potential harms.

Through our analysis, we have distilled an agenda for further research, specifically focused on mechanisms to direct data commodification toward greater inclusivity, social benefit, and protections for the vulnerable. In particular,

- We recognize the promise of data commodification to address data monopolies through the empowerment of small data holders (including individuals). We highlight the critical role of **data pools** (e.g. cooperatives, unions or trusts) for increasing the bargaining power and empowerment of small data holders. But the idea of a data pool is only in its infancy. We need to conduct further research to develop the legal mechanisms and technical infrastructure for these solutions and invest in capacity building and other efforts to ensure they are meaningfully accessible to all.

- We recognize the benefits of **pricing data**, including the use of market mechanisms. It can improve transparency by providing individuals with a comparative measure of the value of their information. It promises to address the trend toward underutilization, while providing a possible mechanism for compensating data labor. But we must develop these tools with their limitations in mind. They are not ideologically neutral. They may be used to exacerbate existing inequalities. Privacy cannot not become a luxury good. And, finally, we need to consider safeguards against excessive self-commodification by exploring limits on the commodification of data.

- We recognize the promise of **blockchain-based data exchanges**. They can facilitate the exchange of data at scale, including from small data holders. They provide an infrastructural underpinning for small data holders to consent to and benefit from the *use* of data (vs. collection). They can support the formation and functioning of data pools. But to ensure that we reach broad participation for broad benefits, which is key to realizing the promise of data commodification, capability building is required to ensure that not only the digitally-advantaged are able to participate and benefit. We also need to clearly

articulate the opportunities and risks that come with decentralized solutions (e.g. no central oversight) since networked structures are not an inherent force for good.

To address concerns about data commodification, it will not be enough to study the mechanisms above. In addition,

- Data commodification is driving toward **propertization of data** (i.e. the idea that 'data is property'). We must define the 'bundle of rights' over data that result from propertization, and how traditional property rights, such as the ability to block the use by others, mesh with the desire to break the trend toward underutilization of data. There is a very real risk of corporate capture of the data commodification movement. Without careful examination of current (or desired) ownership of data, data commodification via propertization can be used to entrench existing data monopolies. Finally, current models of ownership centered on individuals are likely a bad fit for data that is inherently relational (genetic data, transaction data, etc.).

- Data commodification is likely to exacerbate the trend towards commercialization. But it is important that social and humanitarian uses of data are not sidelined. The traditional reliance on data philanthropy and public-private partnerships has been inadequate. The idea of data **as a public good,** and not simply *for* the public good, is a conceptual shift that may be needed to unlock the power of data for social good. We need to concentrate on efforts to safeguard and realize data as digital public goods in the digital commons.

Finally, there is precedent across societies for restricting commodification. It is not morally permissible to commodify human beings, while the commodification of human body parts is permissible but limited and tightly regulated. We have yet to answer these questions for the commodification of data. The trend of data commodification has implications for a range of normative agendas—including human rights, gender equality, inclusion, equity and the well-being of people and the planet. Are there moral redlines that we need to put on the commodification of data or circumstances to safeguard human dignity, autonomy, and flourishing? While these questions may seem theoretical, their answers underpin the responsible evolution of data commodification, and they must be addressed to mitigate its adverse effects.

Jain
Family
Institute

Acknowledgments

We express our gratitude to everyone who participated in the peer exchange on data commodification organized by the Jain Family Institute and co-hosted by UN Global Pulse in January 2022. This report is very much grounded in the expertise shared during the exchange and discussion. To all of you, thank you!

We would also like to thank our colleagues at the Jain Family Institute, in particular, Bakari Sibert, Sanoja Bhaumik, Molly Dektar, Halah Ahmad, Jack Gross, and Liza Paudel, and at UN Global Pulse, in particular, Caroline Alewaerts and Ayham Al Maleh, for support and suggestions.

All opinions expressed in this report belong to the authors. All mistakes are theirs, too.