

# A semi-commons approach to data governance

IT for Change

Aug 2021

# Framework: Data as the new economic object in the circuit of capital

Nature builds no machines, no locomotives, railways, electric telegraphs, self-acting mules etc. These are products of human industry... the power of knowledge, objectified. The development of fixed capital indicates to what degree general social knowledge has become a direct force of production, and to what degree, hence, ***the conditions of the process of social life itself have come under the control of the general intellect and been transformed in accordance with it.***

- excerpt from *The Fragment on Machines* segment of  
*Grundrisse (1857-61)*

# The data wild west as a product of a resource governance crisis

- A 'finders, keepers' scenario exists in data because of which first mover firms enjoy de facto possession rights over the data they collect, through overbroad application of trade secrets protection.
- The hegemonic 'data must flow' illogic in global policy debates has enabled the planetary march of data extractivism, consolidating a new regime of primitive accumulation.
- This resource governance deficit has produced a data wild west, characterised by the twin-ills of misrecognition (identity-based exclusion and harm) and maldistribution (unfair distribution of access and benefits)

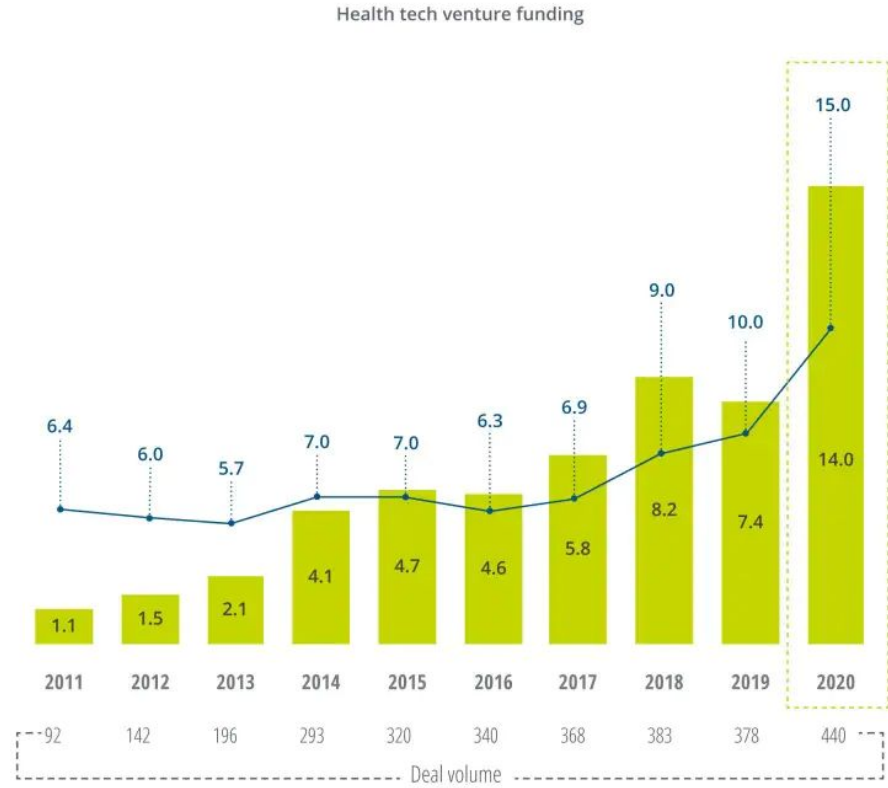
The global healthcare analytics market size is expected to grow from USD 11.461 Billion in 2019 to USD 40.781 Billion by 2025, at a compound annual growth rate (CAGR) of 23.55% (PR Newswire 2020)



FIGURE 1

### Health tech venture funding reached record levels in 2020

■ Overall funding (US\$B)    — Median funding per deal (US\$M)



Source: Deloitte analysis of Rock Health's Digital Health Funding Database.

# Current approaches to data governance: A stocktaking

# The dominant individualist approach

- Individual subjects have quasi-ownership rights in their data, with the right to determine if, and on what terms, their personal data enters the data market through a notice-and-consent regime, within the boundaries for the data market specified by personal data protection legislation (Viljeon 2019)
- Anonymised personal data and machine-observed data that do not have personal identifiers at the point of collection are treated as alienable, non personal data, whose free and unrestricted flows must be maximised

# The dominant individualist approach

- Non-personal data is treated implicitly and automatically as the private property of data processors.
- The question of the economic claims of citizens in the data value generated from *their* anonymised personal data or *their* data footprints in machine-observed data is completely sidestepped.

(Citizens only have the right to be protected from willful/inadvertent deanonymisation in the processing of non-personal data)

# Shortcomings of the individualist approach

- Inattention to privacy risks stemming from downstream uses of machine-observed data points that do not have personal identifiers at point of data collection (Nissenbaum 2019)
- Failure to address economic fairness in data value chains
- Proposed solutions do not go far enough:

Competition law grants defensive and circumstantial, rather than categorical or abstract, access claims, to **address market abuse** (Ulrich 2019). Meaning, within such a framework it is not possible to argue a positive claim to **non-exclusive access** (in data).

Approaches like the EU's 'data producers' right' for NPD (introduced in 2017, ignored in Data Strategy doc 2020) create a complex mesh of private claims to excludability. Multiple claims can prevent appropriate aggregation of a resource, making it difficult to realise its optimal value.



# Data stewardship - an alternative collectivist approach emerging in the global North

- Institutional arrangements where a group of people come together to pool their data and put in place a collective governance process for determining who has access to this data, under what conditions, and to whose benefit (ODI 2019)
- Heuristic adopted across the ideological spectrum to describe a wide range of experiments: some that attempt to actually distribute gains and improve collective bargaining, while others that are just ethics-washing exercises by powerful firms.

## Data stewardship (contd.)

- Data stewardship models are at risk of ending up as a device for large data monopolies to externalise their regulatory burden, reducing administrative costs and reputational risks in the process of data collection and processing (Stuart Mills 2020).
- Even in the best case scenario, data stewardship recasts data into a 'pro-capitalist' commons (George Caffentzis). Data market propositions that convert the social relational resource that is data for capital accumulation leave untouched the unequal ownership of data as a means of production.

# Community data - an alternative collectivist approach emerging in Indian policy circles

- Data resources are seen as akin to natural resources in that both are part of “a nation’s or community’s collective resources as arising from their natural and/or social spaces, and should be governed as such”.
- Community data approach moves the needle in the data governance debate by acknowledging data’s social moorings, and hence, anchoring data rights in its associated communities, and also identifying data-enabled value creation as a vital national public policy issue.

# Community data (contd.)

- Constitutional principles for fair distribution of a community's natural resources are a useful normative compass; but the institutional governance regime for natural resources cannot be replicated exactly in data - an immaterial resource without clear boundaries.
- Not only are we all contributors to multiple data communities at the same time, we may also find ourselves in the target community of data-based businesses even if we are not part of the source community (the group from whom data was initially compiled). This raises complex issues for the rules of exclusion-inclusion and the evolution of representative decision-making mechanisms in the institutional mechanisms of data governance.

Data semi-commons: a new approach to  
govern data for distributive integrity

# The primary problem to be addressed in governing data resources

How can exclusive ownership of data in the hands of a few firms be wrested away for democratising its use and benefit?

The problem in data is one of assembling ownership to a social optimum to address what Heller (2013) terms “wasteful underuse”. Multiple and fragmented parcels of ownership in this situation may lead to a tragedy of the anticommons.

But at the same time, we do not want an unconditional open access regime.

An appropriate resource regime for data must preserve the ‘openness of use’, also promoting ‘accessibility’, that is, the freedom for all economic actors to meaningfully leverage data for unlocking its value (extending the thinking on the intellectual commons about the limits of appropriability)

# The unique nature of data resources

Data is constituted by three distinct layers –

- a) the semantic/content layer, which encapsulates the information being encoded;
- b) the syntactic layer, which is the representation of the information collected as machine-readable datasets; and
- c) the physical layer, the networked infrastructure through which data is extracted (Stepanov 2019).

What we are dealing with in the governance challenge to democratise data value is the ability of the regime to prevent the possessor of the physical-syntactic layers (the network-data architecture) from claiming exclusive rights over the semantic layer in all possible re-uses of the data generated.

Any collectivist framework for data therefore needs to deftly manage the delicate balance between “openness as non-exclusive accessibility” of data’s syntactic content and “openness as duty to nurture use” of data’s semantic propositions.

# A semi-commons regime for data

The semi-commons framework creates and demarcates the boundaries between common property and private property in data resources ownership.

Ownership, in this framework, is understood not as the simple and non-social relationship between a person and a thing, but a complex set of legal relations in which individuals are interdependent and which determines the limits of an individual's or group's freedoms to “use, possess, enjoy or transfer” a particular asset (Johnson 2012)

With a normative baseline in ‘freedom of open use’ in data resources, the semi-commons approach proposes a rights-based resource ownership regime, with varying degrees of differentiated access rights and associated conditionalities for economic actors across the spectrum.



# A semi-commons regime for data (contd.)

## A. Dataholders' right to non-exclusive access

Dataholders have a right to non-exclusive access in the base layer of data that they have collected -- raw, non-personal data, personal data and aggregate data sets combining personal and non-personal data - without exclusive possession rights.

Dataholders have a perennial obligation to respect privacy in the processing of any category of data and have an additional mandatory duty to share data as required by their obligation to respect the rights of data-seekers.

# A semi-commons regime for data (contd.)

## B. Right to seek data

The corollary of the right to non-exclusive access in the data semi-commons is the right to seek data in the datasets collected, aggregated and controlled by for-profit legal entities, altruistic organisations and public agencies through an entitlement of accessibility.

Data seekers can be individual data subjects, public agencies or private legal persons.

Data seekers may access raw non-personal data and/or aggregate non-personal data (except individuals who only have a right to their own data).

# A semi-commons regime for data (contd.)

Table 1 - Differential rights to seek data in the data semi-commons<sup>17</sup>

Data holder →	Private for profit entity		Altruistic data organisation		Public agency	
	Raw non-personal data	Aggregate non-personal data	Raw non-personal data	Aggregate non-personal data	Raw non-personal data	Aggregate non-personal data
Access seeker ↓						
Individual data subject	Right to data access and portability	No	Right to data access and portability	No	Right to data access and portability	No
Private for-profit entity	Right to conditional access	Right to conditional access	No	No	Right to conditional access	Right to conditional access
Altruistic data organisations	Right to conditional access	Right to conditional access	Right to conditional access	Right to conditional access	Right to conditional access	Right to conditional access
Public agency	Authority access	Authority access	Authority access	Authority access	State policy	State policy

# Differential rights to seek data in the semi-commons (contd.)

Data seekers do not necessarily have an unconditional right to access. Access may vary from one scenario to another, and boundaries will be differentially determined through appropriate institutional mechanisms.

Conditional access depends (at minimum) on the following:

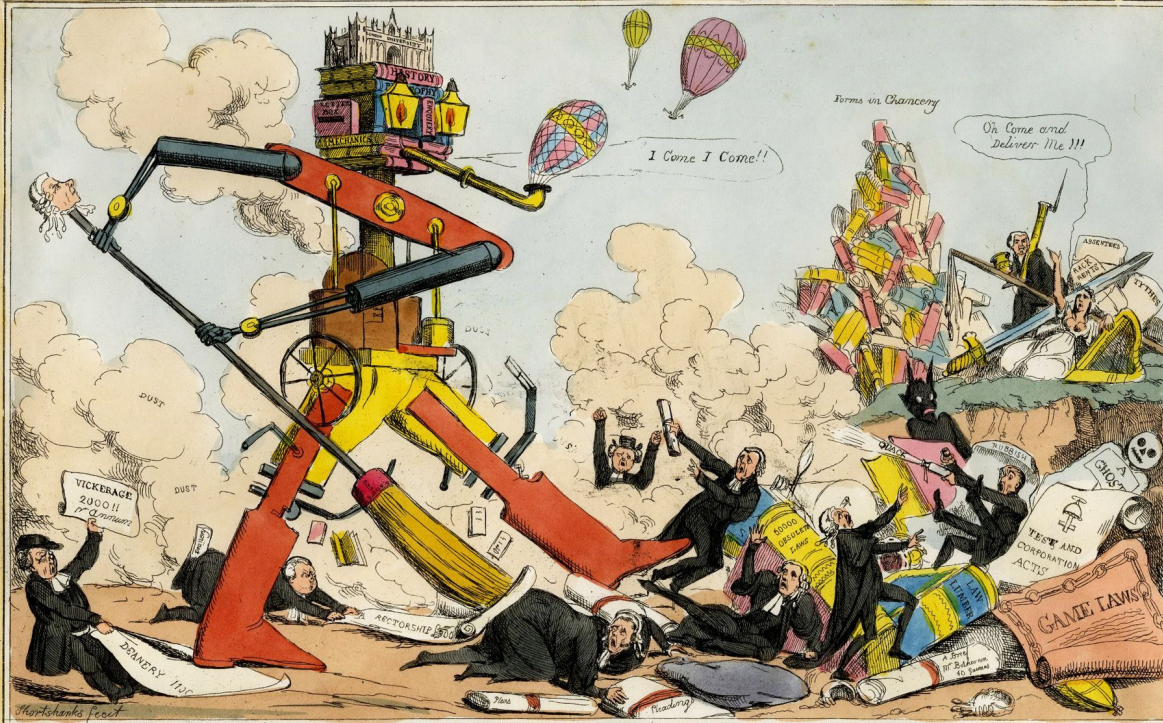
- The type of data seeker and the type of data holder from whom access is sought.
- The specific parts of raw, non-personal data and aggregate non-personal data in which access is sought.
- The purpose (contextual applicability) for which data is sought.

The right to seek data is conceptualised as an entitlement granted through law. It does not, however, preclude rights that arise out of private contracts in the data economy. The establishment of limits of operations of private contracts in the data economy and the downstream rights they produce have an implication for economic fairness. The answers for this cannot be found within the semi-commons itself, and this needs a whole-of-economy approach to governing the data economy.

# Some considerations for institutionalising the semi-commons in the health data domain

- How do we delineate the differential rights to seek data in health data commons across different classes of data seekers, considering that most health data commons are mixed data sets? (where separation into categories of personal data, raw non personal data and aggregate non personal data may not be very easy)
- How important is investment in public health data infrastructure to support the creation of a health data commons? What safeguards do we need (checks and balances that prevent a 'free-for-all' unrestricted access regime) to prevent capture of benefits/data value by powerful players?

THE MARCH OF INTELLECT.



I saw a Vision, a Giant form appeared its eyes were burning lights even of Gays, and on its learned head it bore a Crown of many towers. Its Body was an Engine yea of steam its arms were corn and the legs with which it strode like unto prophesied men called printers use, from whence fell ever and anon small Books that fed the little people of the Earth. It rose and in its hand it took a Broom to sweep the rubbish from the face of the land, the Special pleaders & their wigs also & the Quack Doctors also and the ghosts & those that wore Horns & the Crowns of those that set themselves above the laws & the Delays in Chancery it utterly destroyed, likewise it swept from the Clergy every Placidity. Nevertheless the Lawyers & the Parsons & divers others kicked up a great Dust!!!

Thank you