

Democratizing the information society





AUGUST 2022 Alek Tarkowski, Jan J. Zygmuntowski CHOOSING WHO CAN ACCESS DATA AND USE IT IS A CENTRAL SOCIETAL QUESTION WE HAVE TO ANSWER.

WE NEED TO DEVELOP A SUSTAINABLE INFORMATION ECOSYSTEM THAT SHIFTS THE POWER BALANCE, AND CONTROL OVER DATA, BACK TO SOCIETIES.

THIS CAN BE ACHIEVED THROUGH DEMOCRATIC MANAGEMENT OF DATA AS A COMMONS.

DESIGN OF DATA COMMONS SHOULD BE BASED ON THREE PILLARS:



STEWARDING ACCESS:

achieved by setting rules for managing who gets access, and under what conditions, while preserving rights



COLLECTIVE GOVERNANCE:

ensured through participation and democratic oversight of a trusted community



PUBLIC VALUE:

generated through a clear purpose, capacity-building and sharing

5

- 6 _THE PURPOSE OF THIS PRIMER
- 9 _PATHS TOWARDS DATA COMMONS

12 _DATA AS A RESOURCE AND AS AN INSTITUTION

- **12** *Structuralism: data as a shared resource*
- **14** *Functionalism: data of and for the community*
- **15** *From ownership to access*
- **16** *From individual to collective rights*

19 _THE DATA GOVERNANCE TRILEMMA

22 _DESIGN BLUEPRINT FOR DATA COMMONS

- **23** Stewarding Access
- **24** *Collective Governance*
- **25** *Public Value*

27 _TWO PATHS TO DEMOCRATIZING THE INFORMATION SOCIETY

- 28 _BIBLIOGRAPHY
- 31 _ABOUT

INTRODUCTION

Choosing who can access data and use it – for knowledge-based decision-making, building AI products or conducting research for the public benefit – is a central societal question we have to answer. The challenge is to develop a sustainable information ecosystem that shifts the power balance, and control over data, back to societies, through the democratic management of data as a commons.

Digital data is at the center of many contemporary struggles. We face a Foucaultian reality (Foucault, 1975), where data-driven knowledge produces wealth and power, and respectively the powers that be produce troves of Big Data. This data is not a passive "oil" to be extracted. Rather, the wealth of data is produced by our daily labor and appropriated by increasingly colonizing technologies. These enter ever new spheres of life to capture value, while leaving people with little agency over their governance (Couldry, Mejias, 2019; Zygmuntowski, 2020).

Our sovereignty as individuals and society, our rights and our productive capacities depend on how data is governed: collected, accumulated, shared, accessed, used. And at the same time our economic, legal and cultural systems – the ones that have seemed outdated for a long time already, but also many of the latest innovations and experiments – are ill-prepared to deal with data.

Clearly, the current extractive regime is harmful, and even supposed benefits of innovation and economic development are seriously questioned (Mazzuca-to, 2018). Searching for a solution, some propose to strengthen property rights, assuming that data ownership will allow us to participate as rational agents and benefit from surveillance markets. Others aim to expand data protection and privacy rights, through an approach focused on dignity and minimisation of data use.

Proponents of data commons propose a third approach. One that favors **tapping into collective and democratic governance of data as a resource that is generative and serves to create public value**. The idea of the commons is today gaining visibility and significance as a byword for any approach that opposes power concentration and predatory extraction of data (Marella, 2017). By building data commons, persistent and systemic problems related to the data economy can be addressed.

THE PURPOSE OF THIS PRIMER

With this primer, we want to translate conceptualizations of governance of data as a commons into a framework for designs that bring this idea to life – in particular through public policies.

There is today a wealth of theoretical frameworks that apply the idea of the commons to the data governance debates. We understand commons-based approaches in the broadest sense, encompassing all frameworks that challenge the role of individual property as dominant means of organizing social relationships (Marella 2017). We therefore see it more broadly than the ostromian tradition that treats the commons foremost as a social institution for managing resources by communities (Ostrom 1990).

Against the backdrop of this wealth of theoretical concepts, the number of data commons initiatives is relatively small. Furthermore, concepts of the data commons are not translated into public policies.

The European Strategy for Data, with its focus on ensuring access and use of data, creates an opportunity for introducing commons-based data governance. The policymakers themselves have proposed some policies that support, for example, data cooperatives or limited sharing of private data in the public interest – but the scope of these policies needs to be expanded.

In the first part of this primer, we present key issues and approaches that provide a theoretical underpinning for data commons designs. We begin by conceptualizing data as both a resource and an institution, combining structuralist and functionalist perspectives on data commons. We also highlight two trends in thinking about data and society that are necessary to create data commons: from ownership to access, and from individual to collective rights.

Based on this, in the second part, we propose a set of design principles for building data commons. The design goal is to solve the data governance trilemma: to balance public interest, economic value and fundamental rights. We define nine principles that need to be considered, divided into three pillars: Stewarding Access, Collective Governance and Public Value.

In this primer, we chose not to take Ostrom's design principles for the commons (Ostrom 1990) as a starting point – as many conceptualisation of data commons do. Largely because we believe the policy and practoce communities need to move beyond universal principles and towards a resource-specific operationalization, focused on the characteristics of data.

The general term "data" hides a wealth of resources and information that are contextualized and highly varied, and require different forms of governance. The most important distinction is the one between personal and non-personal data (with continuing debates where the border line can be drawn). But further differences are due to the sectors or spheres of social life that data concerns - health

data is different from shipping industry data, and public statistical data is different from data collected by online platforms about its users.

In this primer, we do not consider these distinctions, although we acknowledge them. We aim to propose principles that are general enough to leave room for considering this diversity, and which should apply to all data commons initiatives. Hopefully, these general rules will serve as a basis for more specific blueprints and frameworks for different types of data.

We are in particular interested in public policies that can bring data commons into life. More generally, there is a need to develop alternative data governance logics that oppose dominant data appropriation logics (Bodó et al, 2021). And public policies will play a key role in ensuring that these alternatives are developed, and then made sustainable.

We hope that this primer will help to provide much needed alignment for data commons advocates. In doing this, we are building on the work of multiple data commons theorists, activists and advocates.



PATHS TOWARDS DATA COMMONS

PATHS TOWARDS DATA COMMONS

Data commons are digital data that are collectively stewarded and governed by a community. Data commons are not data alone, but refer to the entire systems, including institutions that enable sharing and ensure public value generation and the community that is involved in commoning (de Angelis, 2017).

The practice of commoning with regard to data comes in many flavors. The differences are due not only to the specificity of data that is being governed, but also due to institutional design: legal frameworks that are employed, type of actors constituting the commons, or its recognition by public authorities.

The terminology used is often overlapping, with similar terms – including "data commons" – used to describe different models; and similar models defined using different terms. There is no consensus on a single definition of not only the data commons, but some of the specific forms as well.

Policy primitives for data commons

The key proposed terms, or forms, can be understood as "policy primitives", with which different forms of data commons can be designed and built:

_ OPEN ACCESS COMMONS, the least restricted repositories of easily shareable sets of data, where monitoring access is not necessary,

_ DATA COLLABORATIVES, voluntary public-private agreements to share data to achieve synergies along the lines of "data for good",

_ DATA COOPERATIVES, grassroots initiatives employing the democratic model of cooperativism and social entrepreneurship to govern data of its members,

_ PUBLIC DATA COMMONS, institutions such as agencies, banks and trusts established to provide systemic solution of data governance in the public interest,

_ DATA TRUSTs, authorised third-party managers of data, bound by fiduciary obligations to act in the best interest of beneficiaries,

_ DATA UNIONS, collective bargaining institution, distributing revenue for members and assisting them in data strike actions,

_ COMMON DATA SPACES & DATA ACCESS BODIES, connecting sectoral, interoperable frameworks of standards with designated authorities (as seen in European Health Data Space proposal).

While the variety of concepts is proof of ongoing debate, there is a growing need to discern and evaluate them in terms of their fit for purpose: their potential for

generative productivity, data justice or structural change-making. This assessment will be largely contextual, as different institutional designs are needed for different types of data and different social context.

Ultimately, data commons designs should have the capacity to reshape and change the power balance in data-driven economy and society.

While academic literature focuses on comparing these approaches and highlighting their context-dependence, data commons advocates often engage in debates that aim to prove the advantage of one approach over others. The strongest debate pitches advocates of bottom-up, or grassroots, initiatives like the data cooperatives against proponents of top-down or "statist" approaches. **We see these initiatives as fellow travelers, engaged in a productive dynamic that serves to enrich the data commons movement.**

In fact, data cooperatives would benefit from scaling via becoming public infrastructures, or even being nested within existing welfare systems. And in turn, public data commons and data access bodies could gain by embracing less bureaucratic management and providing means for participatory governance. And regardless of the type of data commons design that is deployed, the same elements need to be included, designed and built.



DATA AS A RESOURCE AND AS AN INSTITUTION

_DATA AS A RESOURCE AND AS AN INSTITUTION

The idea of the data commons has been developed in parallel from two differenet perspectives. It connects two main lines of thinking, visible both in scholarship and in practice, related to a sustainable, inclusive society that is data-driven:

1. DATA STRUCTURALISM, which holds that data as a type of resource have inherent properties that require collective governance models for their proper functioning;

2. DATA FUNCTIONALISM, which proposes establishment of collective rights in data to address social injustice and empower communities to whom the data pertains (Fia, 2021).¹

The former draws from classic works on natural common-pool resources and their sustainability, and from a strong foundation in economic thought. The latter is rooted in legal studies, stressing the social dimension of governance and ful-fillment of fundamental rights. **Effectively, data structuralists claim that data** <u>are</u> **commons, whereas data functionalists claim that data** <u>should be</u> **commons.**

These two **approaches complement each other, and they should be connected.** They are the equivalent of the wave-particle duality in quantum physics. Sometimes one concept illuminates the behavior of a quantum object better than the other, but separately, neither wave or particle describe it completely. The structuralist and the functionalist account both provide strong, complementary input into the data governance debate.

Two different views of one phenomenon allow us to grasp it fully. And the dvocates of data commons need to make both arguments: that by treating data as a commons we can fully benefit from inherent properties of data that can be limited or blocked; and that we can design and establish institutions that will allow us to even better use data to attain societal goals.

Structuralism: data as a shared resource

The structuralist approach evaluates data from a systemic perspective, asking what is the best resource management strategy for data production and use, given the costs and benefits. It assumes that understanding the nature of data allows us to create a sustainable system, one which generates value while regenerating its socio-economic conditions, by deciding where the benefits and disbenefits fall,

1 There might be data structuralists who claim that data "by nature" should be a free-flowing commodity; and consequently, there might be data functionalists caring about new rights for tech companies, not communities. The terms denote only a perspective, an intellectual focus on either the structural or functional properties of data.

progressively and towards justice.²

Institutional economics – one of the key structuralist approaches – defines common goods³ by their properties: low excludability (difficulty in limiting consumption) and high subtractability (rival consumption threatens them) (Ostrom, 1990). Once these categories are applied to data, the structuralists argue, its characteristics as a commons become clear.

The early-internet era assumed that data are public goods, with non-rival, near-zero marginal costs of copying and overall "intangible abundance" (Biga, 2021). But the reality of the data-driven world is much more complex. Just as clean air, once assumed to be a public good, is increasingly treated as a commons that needs collective governance and protection from over-use, so are data.

Over time, data turned out to be substractable, as **negative externalities diminished the general sustainability of the data ecosystem** (Zygmuntowski et al., 2021). Decreasing trust in digital systems, such as concerns over breaches of privacy, and hesitation towards data sharing stems from this unsustainable use of personal data. As unwanted actors get hold of our data, we become more sensitive to sharing and mistrustful towards not just harmful, but also socially beneficial digitalization.

When it comes to data's low excludability, the cost of limiting data sharing is relatively high - it requires large investments in gatekeeping, meaning building a legal and technological context where data can be privatized. But the hidden alternative cost of not having other societal actors reuse data, thus boosting data-ta-driven innovation, is even greater.

When we face large platforms excluding communities and individuals from access to their data, we are not witnessing an inescapable property of data but an extensive effort to enclose the data commons. Thus, **a structural intervention becomes needed to maximize social welfare** by bringing down the costs of monopolizing data.

There are cases when data functions as a public good – this is true for example for statistical data. In this case, the Open Data model can be applied and the data can be treated as an Open Access commons (Bodo, 2019). Openness as a particular form of access to data should be seen as lying on a spectrum of possible decisions on data governance, together with varied forms of permissioned access. The important shift that is needed – and occurring – today concerns optimizing outcomes of data governance by balancing the problems generated by some sharing (subtractability) with the problems generated by not sharing (excludability).

² We are grateful to Jeni Tennison for proposing this formulation, which reframes what is the typical framing of the cost /benefit analysis of the data commons.

³ The classic work by Elinor Ostrom, which awarded her a Sveriges Riksbank "Nobel" prize in economics, focused on natural resources like fisheries, forests, pastures and ground waters. These "common-pool resources" differ from public goods in that they are rivalrous in consumption and have to be governed to avoid depletion. Public goods are also shared and non-excludable, but hardly rival; national defense is a typically given example.

This turns the attention from considerations on the extent of sharing - towards the conditions of this sharing.

Given that data can be used by various actors as inputs to create private, social and public value, they can also be viewed as an infrastructural resource. Similar to utilities like energy or roads, data are general-purpose inputs used to produce final use outputs, such as sense-making analyses, AI products, or render services more accurately. And infrastructural resources are "means for many ends", and thus have to be treated like shared resources (Frischmann, 2012). As a result, the commons becomes a resource management strategy for administering open or controlled access to the infrastructure.

Governance of data as a commons is particularly relevant for non-personal data, which is often infrastructural, in the sense described above. This means that it can functionally support the creation of other goods and services. Governance of such immaterial infrastructure is based on open access to resources, non-discrimination and stewardship of data for the benefit of society as a whole. Therefore, even for very large social institutions like the state, for which it is often hard to conceptualize the community, there is a relation between society's members and the data, and a collective dimension to ownership that should be acknowledged.

Functionalism: data of and for the community

The functionalist approach observes that data is always produced in complex systems, in which social actors interact with technological and legal tools (Kitchin & Lauriault, 2014). The laws and governance mechanisms that we create always impact who controls the resource and benefits from it. Thus, **data commons are functional: they ensure enjoyment of fundamental rights and fulfillment of public interest goasls.**

Data can be understood as relational in two ways. On one hand, much of data represents relations between people, and is furthermore structured in a way that can relate people to each other. Almost all personal data is potentially relational (Viljoen, 2021). And almost all data is relational in the sense that it relates to people that are involved in its production (Marella 2016).

The two meanings of data's relational character are furthermore connected with each other. **The more data is relational in character, the more collective rights in data should play a role, and the more it is the community that should have rights in data.** (Sadowski et al., 2021). This type of a collective rights perspective can be employed in varied contexts: for example, to justify the need to support institutional forms that help express collective rights in data; or for the need to ensure access to data aggregated by digital platforms. A key challenge in this regard is defining the appropriate community (or other collective actor) in whom the collective rights are vested, and then creating an appropriate governance framework.

Governance of data is therefore not merely a technical, but also a social and political question. Data is a "democratic medium" (Viljoen, 2021) capable of strengthening a community by re-routing patterns of usage and value creation. Our ability to claim control over data determines the extent to which we are either an object of surveillance, or a subject of data-driven innovation. And this ability increases when people have the capacity to act together and collectively manage data.

And on the contrary, collective data rights can leverage the aggregated value of data towards the aims chosen by a community. These communities vary immensely: they could be Indigenous people of the same nation, rare disease patients, citizens of the same municipality or platform workers.

This approach to data commons ensures that data is protected from inappropriate use, but more importantly is used according to the needs of the community. Hence the necessity to establish new governance models and institutions that address the particularities of managing data commons (Prainsack, 2019).

From ownership to access

Digital data can, by virtue of its characteristics studied by structuralists, be a widely - almost universally - available resource. Yet in reality the ability to use data, and also to control its use is limited to a limited number of actors. There is a discrepancy between the material, objective costs, benefits and possibilities related to data as intangible productive force - and the historically emergent mode of production shaped by capitalist extractivism of data. The former is a binding set of possibilities, but the latter is only one specific mode.

The dominant mode of control of data in today's data economy is the generation and appropriation of large-scale datasets. And as digital power becomes increasingly centralized, so does control of data. This is based on three types of restrictions to data access: technical measures, exercise of economic power, and legal tools - intellectual property rights in particular.

There is a group of critics of the currently dominant data regime who claim that the best solution is based on decentralization combined with private ownership of data and thus monetization of data use. It supposedly would ensure a fair pay for labor that we all conduct when producing data. This would further serve to bolster self-sovereignty of citizens, as they manage their own data for their personal benefit (Arieta-Ibarra et al., 2018).

> Yet monetized data ownership will on its own not help to avoid the problem of data commodification, even if managed through decentralized data intermediaries (Käll, 2020). This is because selling our data leaves the current data value cycle untouched.

Data governance based only on private ownership, when combined with network effects inherent to digital communication networks, will lead to a "winner takes all" scenario (Acemoglu et al., 2019). We might get a piece of the cake, but a piece of one that will nevertheless be produced by Big Tech and surveillance industries.

From a data ownership perspective, access to data is just an exception to the general rule of owning data as property. From the perspective of the data commons, community-governed access - and the ability to exclude from access - is the rule. Such access can be narrow and limited to members of the community, or broad – and often also stretching the concept of the community, as is the case with Open Access commons.

Concern with access shifts focus to the rights and needs of those aiming to use and benefit from data, rather than just of those who "have" the data. **Asking who "owns" data is much less productive than asking who defines the rules concerning data, meaning: who can control, access, use and re-use data, and then determine who else can do the same.**

The concept of the data commons also shifts the debate to generative functions of data as a resource. It thus becomes focused on ensuring a positive freedom "to operate, instead of power to appropriate" (Fia, 2021). **Choosing who can access data and use it for knowledge-based decision-making, building Al products or conducting research for the public benefit is a central societal question we have to answer, instead of relying on individual property rights to solve this question for us.** In other words, we as members of societies need to answer it ourselves, and not depend on the market or the state to do it for us, and without us.

From individual to collective rights

The introduction – through the GDPR regulation – individual rights of data subjects to control access, rectify or even delete data was a major regulatory win of privacy advocates in the EU. It is as if we finally grasped – as a society – the inherent link between a person and the data connected to that person. This is today the principal contribution of data functionalists to developing rules and social norms for data. But ensuring and protecting fundamental individual rights in a digital world is not enough.

As much as the GDPR was needed, it (and other similar regulations) frames management of own data as an individual right. Similarly, harms are framed largely in terms of individual privacy. Upcoming EU regulations, such as the AI Act and the Data Act recognize that some harms occur strictly in the societal macroscale. And it is becoming clear clear that **pushing the responsibility of managing own data on people leads to "cookie (banners) consent"**⁴, and leads to individualizing further systemic issues, and harms in particular (Janssen et al., 2020).

Nevertheless, some advocate for even more individualized rights as a solution. For example, Shoshana Zuboff proposes the "right to sanctuary", meaning a title

⁴ Cookie banners are the widely used privacy notices in the forms of pop-ups asking about consent for tracking on a website. Most people never read what the banners say, are unaware that large Consent Management Platform companies abuse consent and are coerced into accepting all trackers by dark patterns (opaque, time-consuming design). What we call here "cookie consent" means a formally stated, yet in fact forced and illegal consent (see: Belgian DPA v IAB Europe).

to an (individual) space free from surveillance (Zuboff, 2018). Similarly, advocates of data intermediation based on the idea of data ownership also shift the focus of governance, and thus responsibility, to individuals.

Yet individual rights and personal choice are insufficient in face of problems of collective nature, largely due to the relational character of data (IT for change, 2021). This is best illustrated by data portability rules in the GDPR, which are commonly seen as a failed attempt to ensure greater data access through regulation based on individual decisions to move data.

Just as we established new institutions and democratic methods for all levels of societal ordering, including the right to nations' self-determination, we now need to embed collective data rights in the participatory decision-making over data. State-level actors can play a key role, but other actors can engage as well, at different scales and layers of the society. This will also help bring about the autonomy and digital sovereignty that communities lack so far. And collective rights will benefit market actors, trusted by the community to be granted access. Data sharing mandates, such as the ones currently proposed in the EU Data Act, are an opening for transfer of power away from monopolized data silos (Tarkowski et al., 2022).

As a collective, people can get better terms for their shared interest. A great example is collective bargaining, which allows groups of workers (e.g. represented by a trade union) to negotiate better terms with the employer. Another example are class action lawsuits, which award all class members despite only some of them battled in court⁵. In fact, Google is facing a class action-style lawsuit in the UK over an unlawful use of medical data gathered by DeepMind from the Royal Free London NHS Foundation Trust.⁶ And in another case, Google and Facebook face class action suits in Illinois, forcing the companies to pay out damages to users, and to introduce limitations on some of their facial recognition tools.⁷

A data governance framework therefore needs to shift its basis from individualist data subject rights and their enforcement to institutional forms that can ensure collective governance. This shift from individual to collective control and responsibility is necessary due to the characteristics of Big Data, and in order to curb the resulting power imbalances (McMahon, Buyx and Prainsack 2019). They will also help to develop our capacity to manage data as a productive, socially beneficial resource.

⁵ Class action lawsuits are a model of group litigation developed in the common law, especially in the United States, and to some extent adopted in other countries also with civil law systems. The idea is that when a harm has been done to a collective, large group, instead of multiple cases a single case is proceeded where the representative sues on behalf of the class. The class is usually constructed depending on the harm - it could comprise consumers of a given company, workers or patients. Examples of class action include litigation against pharmaceutical companies, banks and retail corporations.

⁶ https://techcrunch.com/2022/05/16/google-deepmind-nhs-misuse-of-private-data-lawsuit/

⁷ https://www.theverge.com/2022/6/6/23156198/google-class-action-face-grouping-biometric-information-illinois-privacy-act



THE DATA GOVERNANCE TRILEMMA

THE DATA GOVERNANCE TRILEMMA

Data commons present a way out of the trilemma: how do we balance protection of fundamental rights (both individual and collective), generation of economic value and the public interest in access and use of data?

We argue that data commons approaches are an effective means to balancing the three different goals. We are not arguing the opposite, that effectively addressing these three goals at once always leads to the creation of a data commons - it is a necessary, but not a sufficient step. Ultimately, it is the collective and democratic character of data commons that makes them unique.



There is an alignment between the commons and fundamental rights (which is highlighted by the functionalist approach), as commoning of data and protection of rights can be mutually reinforcing in the fight against surveillance and extraction for profit. At the same time, paying too much attention to data rights (especially in their individual aspect) slows down the development of socially valuable forms of data use. A classical example of this are barriers to medical research due to overarching privacy concerns, which commons-based approaches try to solve.

But data commons also include a concern for shared societal welfare (stressed by the structuralist approach). Because the question that begs to be asked is not just what private entities may not do with data, but what data can do as a shared resource for citizens and communities. Beyond the management of data for economic growth and private benefit, and the protection of human dignity and rights in the context of data usage, lies the space of accessing data to develop better public services, improve decision-making and empower citizens. In other words, there are public interest considerations that should be balanced against economic and dignitarian goals.

The compatibility of the commons and the public interest is regarded as contentious - and for a reason. The creation of modern states and universalizing concepts of publics displaced in modern times legacy forms of governing shared resources - such as traditional forms of commoning. But we increasingly observe the emergence of public-common partnerships, which reimagine the role of state institutions and foreground a politics of the commons, especially in the urban context (Russell et al., 2022). The declaration of the French Presidency of the Council of the EU calling for "a new initiative for digital commons in Europe" (EU Council, 2022) is yet another example, on an international scale⁸. **The public is reinvigorated by the commons, whereas the commons achieve scale and systemic impact.**

Instead of shunning from the conflict between those goals, or enshrining the "right" choice in law, data commons hands over decision-making to the community. In other words, data commons can maximize the benefits while minimizing the negative outcomes, by deciding where the benefits and disbenefits fall, progressively and towards justice.

By leveraging collective control and governance of access to data by a community, data commons improve the act of balancing the values and goals entrenched in the trilemma.

8 Recommended read on the French Presidency declaration and the Digital Commons: https:// openfuture.eu/blog/vive-les-biens-communs-numeriques/

4

DESIGN BLUEPRINT FOR DATA COMMONS

_DESIGN BLUEPRINT FOR DATA COMMONS

For data to be governed as the commons, we need an appropriate design of the data ecosystem. Defining the basic elements of a design blueprint for data commons is necessary for a plurality of institutions, initiatives and infrastructures to work together, or at least in parallel, on attaining shared goals.

Design of data commons needs to consider three pillars.

_ STEWARDING ACCESS: achieved by setting rules for managing who gets access, and under what conditions, while preserving rights,

_ COLLECTIVE GOVERNANCE: ensured through participation and democratic oversight of a trusted community,

_ PUBLIC VALUE: generated through a clear purpose, capacity-building and sharing



Stewarding Access

Data commons deploy various forms of access to ensure on one hand that generative characteristics of data as a resource are not limited, and on the other that it is shared in a way that is sustainable, preserves rights and minimizes risks. Design decisions that concern Stewarding Access establish the rules and means for deciding who gets to access data and under what conditions. There is a tension here that needs to be resolved: between Open Access commons and stronger, permissioned forms that limit access through more refined governance. Stewardship also entails maintenance of the data and related infrastructures – as ensuring access requires large amounts of effort to collect, store and maintain quality of data.

__SHARING FRAMEWORK – There can be no data commons without accessible data. Therefore legal frameworks and tools are necessary to first make the data available as data commons, and then to allow access and downstream uses. B2B, B2C and B2G data sharing or data altruism are some of the frameworks that are available. Decisions also concern enabling continuous access. Access may require accepting a specific license that the community adopted for data use.

_ PERMISSION INTERFACE – Where Open Access data commons permit everyone to access and use data, other forms of commons need to be based on permissioned access. Thus, a permission interface needs to be designed. The interface may monitor, register and assess impact of requests to access data. Ensuring that the identity of an actor that is requesting access to data is transparent allows for greater accountability, also in terms of preventing harm and levying sanctions if data commons have been abused.

_____PRIVACY-ENHANCING TECHNOLOGIES – Satisfying data protection by design (GDPR) for personal data can be achieved by conscious architecture choice. Since it is the societal objective that is important, not technological novelty in itself, greater protection of rights should be achieved with privacy enhancing technologies (PETs) such as Open Algorithms (which "move algorithm to data"), federated learning, pseudonimization, distributed vetting ledger and others. While the focus of the data governance debate is on privacy, care should be taken to preserve and enhance other rights as well.

Collective Governance

Data commons are linked to the community which manages them, and in many cases generates the data as well. Any other arrangement would constitute an appropriation of the resource, and disempowerment of the people. To establish collective governance over data, there must be either an existing or a newly established entity that can become trusted institutional vehicle for data commons.

_____DEFINED COMMUNITY – In order to ensure democratic governance, the community that is the primary holder of rights in data needs to be defined. In this way, collective rights in data can also be better assigned and represented. Yet this is often challenging with regard to digital data, as traditional community or group formation frames do not apply. The challenge lies as much in conceptualizing the community, as in defining the right institutional level of civic life at which the collective interests should coalesce.

_ TRUSTED INSTITUTION – A trusted institutional actor capable of stewarding the commons is a necessary element of data governance design. Data commons institutions are needed due to limitations of both grassroots organizing and market incentives. Institutionalizing the data commons, and thus supporting them with dedicated infrastructure, funding and capacity, renders them independent from market or state pressures.

_ DEMOCRATIC CONTROL – For the community to have greater autonomy, it has to be directly involved in decision-making. Different forms of democratic participation or accountability can be deployed, including supervisory councils, citizen panels and assemblies, sortition and quadratic voting. Different forms of democratic control can be deployed at all levels of social life, from the local and municipal level to the governance of national datasets.

Public Value

A successful data commons strategy needs to take into account not just the management and provision of data, but also the need to ensure that the generated data-based products and services increase the common good. The notion of public value is useful to emphasize concrete, observable benefits produced for the society as a whole, and not just for the community that manages a data commons. By providing public value, data commons can restructure the data value cycle, change the balance of power and introduce a regenerative function to the data ecosystem. A public value perspective also pays attention to positive externalities of data commons, such as increased data literacy or experiences with civic partic-ipation.

_____MISSION-ORIENTED DATA COMMONS - Data commons initiatives should be guided by the values upheld by the community and oriented towards societal goals. Thus, access to data is not a goal in itself, but should lead to socially beneficial uses. A mission-oriented approach ensures that data commons benefit the society in an egalitarian, inclusive manner, for example by prioritizing or incentivising data use for socially important aims.

_ COMMON WEALTH LICENSING - There is a need to build a new generation of licensing tools that allow access and use rights to be managed in as standardized way as possible. As a general principle, a license for data access and use should aim to build the shared wealth of community, by sharing the products and revenues back with the commons, and with the society – instead of just producing commercial value.

_ DATA LITERACY - All commons have to remain sustainable by not only regenerating their stock, but also the capacity of the community to continue commoning. In the case of data commons, this means supporting projects of redistributive justice and reducing inequalities in the capacity to obtain value from data commons. Broadly understood data literacy includes not just individual education and training, but increases in the capacity of different actors, institutions and communities to make beneficial uses of data. 5

TWO PATHS TO DEMOCRATIZING THE INFORMATION SOCIETY

_TWO PATHS TO DEMOCRATIZING THE INFORMATION SOCIETY

Ensuring sustainable and democratic data governance is one of the main struggles that information societies face. We have already gone a long way from accepting data monopolization as a legitimate business practice. European policymakers are deploying new legislation that seeks to reclaim the power balance and to address societal harms, and pilot projects to develop alternative data governance models are being launched.

But as much as efforts across the world to create robust data sharing schemes are inspiring, hardly anyone has solved the puzzle yet. Too many ideas remain just concepts on paper or on screens, or minimally viable projects that urgently need to scale.

For this reason, we believe that data commons will not be established without a public intervention, at either national or international (e.g. European) level. Without strong policies and regulation to reinforce them, data commons initiatives will either fail or remain marginal in an adverse economic and legal environment. This is due to the fact that the lock-in of the current data value cycle is too strong, and constantly reinforced.

The growth of Open Data – and in particular its significance during the COVID-19 pandemic – has shown the essential role that public actors play in establishing spaces that break away from extractive, property based models of managing data. Public support and funding is also needed for grassroots efforts, like data cooperatives, to scale into viable alternatives that can have a significant share in the data economy. The aforementioned public-common partnerships pave one such way.

Advocates for data commons, in the civic and public sectors, should share a strategy based on two parallel approaches. On one hand, we need to build public data commons institutions, new collective governance entities which can be the cornerstone of the co-creation of public value outside the state. On the other hand, the space of grassroots data commons needs to be resourced and sustained to create safe space for experimentation and creation of new governance models. Both will serve to create more just and democratic forms of governing our data-About Open Future



BIBLIOGRAPHY

- Acemoglu, D., Makhdoumi, A., Malekian, A., & Ozdaglar, A. (2019). Too Much Data: Prices and Inefficiencies in Data Markets. *NBER Working Paper*. <u>https://doi.org/10.3386/w26296</u>
- Biga, B. (2021). The economics of intellectual property and openness: The tragedy of intangible abundance. *The Economics of Intellectual Property and Openness: The Tragedy of Intangible Abundance*, 1–114. <u>https://doi.org/10.4324/9781003098386/ECONOMICS-INTELLECTUAL-PROPERTY-OPEN-NESS-BART</u>
- Bodó, B. (2020). Was the Open Knowledge Commons Idea a Curse in Disguise? Towards Sovereign Institutions of Knowledge. SSRN Electronic Journal. <u>https://doi.org/10.2139/ssrn.3502119</u>
- Bodó, B., Irion, K., Janssen, H., & Giannopoulou, A. (2021). Personal data ordering in context: the interaction of meso-level data governance regimes with macro frameworks. Internet Policy Review, 10(3). <u>https://doi.org/10.14763/2021.3.1581</u>
- Couldry, N., & Mejias, U. A. (2019). *The costs of Connection: How data is colonizing human life and appropriating it for capitalism*. Stanford University Press.
- De Angelis, M. (2017). *Omnia Sunt Communia: On the Commons and the Transformation to Postcapitalism*. Zed Books.
- EU Council (2022). Declaration by the Presidency of the Council of the European Union calling for a European Initiative for Digital Commons. Published 7.02.2022.
- Fia, T. (2020). An Alternative to Data Ownership: Managing Access to Non-Personal Data through the Commons. *SSRN Electronic Journal*. <u>https://doi.org/10.2139/</u> SSRN.3698914

Foucault, M. (1977). Discipline and Punish. The Birth of the Prison. Pantheon Books.

Frischmann, B. (2012). *Infrastructure: The social value of shared resources*. Oxford Univesity Press.

- Janssen, H., Cobbe, J., & Singh, J. (2020). Personal information management systems: A user-centric privacy Utopia? *Internet Policy Review*, 9(4), 1–25. <u>https://doi.org/10.14763/2020.4.1536</u>
- Käll, J. (2020). The materiality of data as property. *Harvard International Law Journal*, 61. https://harvardilj.org/2020/04/the-materiality-of-data-as-property/
- Kitchin, R., & Lauriault, T. (2014). *Towards Critical Data Studies: Charting and Unpacking Data Assemblages and Their Work*. <u>https://papers.ssrn.com/sol3/</u> papers.cfm?abstract_id=2474112
- Marella, M. R. (2016). The Commons as a Legal Concept. *Law and Critique 2016* 28:1, 28(1), 61–86. <u>https://doi.org/10.1007/S10978-016-9193-0</u>
- Mazzucato, M. (2019). *The Value of Everything: Making and Taking in the Global Economy*. Penguin.
- Morton, T. (2013). *Hyperobjects: Philosophy and Ecology after the End of the World*. University of Minnesota Press.
- Ostrom, E. (1990). *Governing the commons: the evolution of institutions for collective action*. Cambridge University Press.
- Prainsack, B. (2019). Logged out: Ownership, exclusion and public value in the digital data and information commons. *Big Data & Society*, 6(1), 205395171982977. <u>https://doi.org/10.1177/2053951719829773</u>
- Russell, B., Milburn, K., Uk, R.-L.-S., & Heron, K. (2022). Strategies for a new municipalism: Public-common partnerships against the new enclosures: https://Doi.Org/10.1177/00420980221094700,004209802210947.
- Sadowski, J., Viljoen, S., & Whittaker, M. (2021). Everyone should decide how their digital data are used – not just tech companies. *Nature 2021* 595:7866, 595(7866), 169–171. <u>https://doi.org/10.1038/d41586-021-</u> 01812-3
- Tarkowski, A., Vogelezang, F., Keller, P. (2022). Public data commons. A public interest framework for Business-to-Government data sharing in the Data Act. Open Future Foundation.

Viljoen, S. (2020). A Relational Theory of Data Governance. SSRN Electronic Journal. <u>https://doi.org/10.2139/SSRN.3727562</u>

Zuboff, S. (2019). *The Age of Surveillance Capitalism: The Fight for Freedom and Power in the Age of Surveillance Capitalism*. Public Affairs New York.

Zygmuntowski, J. J. (2020). *Kapitalizm sieci*. Roz:Ruch Publisher.

Zygmuntowski, J. J., Zoboli, L., & Nemitz, P. F. (2021). Embedding European values in data governance: a case for public data commons. *Internet Policy Review*, *10*(3). https://doi.org/10.14763/2021.3.1572

ABOUT

<u>Open Future</u> is a European think tank that develops new approaches to an open internet that maximize societal benefits of shared data, knowledge and culture.

Authors of this report

<u>Alek Tarkowski</u> is the Strategy Director at Open Future. He holds a PhD in sociology from the Polish Academy of Science. He has over 15 years of experience with public interest advocacy, movement building and research into the intersection of society, culture and digital technologies.

Jan J. Zygmuntowski is a Data Commons Fellow at Open Future. His research project is focused on data commons governance. Jan is an economist interested in development studies, the political economy of technology and the digital economy.

The authors want to thank Balazs Bodo, Tommaso Fia, Jack Hardinges, Paul Keller, Gijs van Maanen, Joe Massey, Barbara Prainsack, Nadya Purtova, Jeni Tennison and Francesco Vogelezang for their feedback, which helped to shape this primer.



ALEK TARKOWSKI



JAN J. ZYGMUNTOWSKI (c) by: Kaja Róg / Silver Second Studio

KEEP UT TO DATE AND SUBSCRIBE TO OUR NEWSLETTER

SUBSCRIBE



HELLO@OPENFUTURE.EU

