

WORKINGPAPER

November 2023

Public Data Commons as a Policy Framework for the Data Impact Hub and the Mutual Commitment Framework

Dr Alek Tarkowski¹

UNU-CPR Working Papers are research papers that have not been peer-reviewed or undergone a thorough editing and publication process. Written by subject matter experts, they offer unique insights and perspectives in response to current debates on issues of strategic interest to UNU-CPR audiences.

¹ Director of Strategy at Open Future Foundation.

Introduction

This brief outlines how a framework for data sharing within multilateral institutions, consisting of a Data Impact Hub and a Mutual Commitment Framework, could be established, based on the principles of commons-based data governance.

It builds on the concept of Public Data Commons, a public interest data sharing model developed by the Open Future Foundation during the policy debate on the European Data Act. This legal act is one of several prior cases of legislation that establishes regulatory frameworks for Business-to-Government (B2G) data sharing. The Public Data Commons framework is particularly suited to enable sharing of data from various sources, and in particular, to secure greater availability of data held by private companies.

Since this type of data will be crucial for the mission of the Data Impact Hub, the Public Data Commons model offers a valuable reference point for establishing the Hub and the Mutual Commitment Framework. A commons-based approach offers a set of principles and more specific policy design guidelines that can help to share data as a digital public goods, enable collective decision making, and balance the need to share with care for data rights and to secure public interest.

The proposed Data Impact Hub in the report of the High-Level Advisory Board on Effective Multilateralism correctly frames the need for pre-positioned data capacities and pre-agreed frameworks if global emergencies are to be properly addressed. The Public Data Commons offers a model in which a trusted institution acting in the public interest not only pre-positions capacities but, over time, builds a body of data that can be shared securely and in a trustworthy manner. This is particularly important in times of “poly-crisis,” when emergencies cannot always be framed as discreet events but have a systemic character and play out over longer periods. For this reason, ad hoc data sharing solutions in the face of emergencies will not be sufficient.

The first part of this brief outlines the Public Data Commons concept as a framework for business-to-government (B2G) data sharing. The second part considers what lessons could be learned from the European debate on B2G data sharing. It offers additional considerations for introducing a commons-based approach for the Data Impact Hub.

Public Data Commons as a Design Blueprint for Data Sharing

Public Data Commons is a data sharing framework that aims to secure public interest goals through commons-based data governance.¹ As such, it seeks to shift the policy debate away from a dominant perspective that treats data as a purely economic asset, available primarily to private companies for financial gain. In this perspective, even data sharing can achieve a competitive

¹ Alek Tarkowski, “Public Data Commons: A Public Interest Framework for B2G Data Sharing in the Data Act, Open Future, 24 May 2022, <https://openfuture.eu/publication/public-data-commons/>.

advantage. A commons-based approach assumes additional value to be gained by non-market use of data, especially since according to researchers, most of the data is never fully utilized.²

The Public Data Commons model also aligns well with the current digital strategy of the United Nations, and most importantly, with the commitment to steward Digital Public Goods, expressed in *The Age of Digital Interdependence* report, and implemented through the current UN roadmap on digital cooperation and the *UN 2.0* policy brief.³

There is currently an asymmetry in the capacity to access and utilize data between market and non-market actors, with a narrow group of prominent companies and online platforms having entrenched positions with regards to data ownership.⁴ Data from the Global Data Barometer⁵ shows that this asymmetry affects, in particular, public bodies around the world, as the mean country score is only 34 points out of 100 on a comprehensive benchmark for public data ecosystems.

The Covid-19 pandemic catalyzed some data sharing efforts, both in the shape of interventions by public bodies and initiatives by businesses that can be framed as “data philanthropy.” Yet these have been limited in scope rather than systemic and have largely focused on the field of health data. The current debate on generative AI models, where the largest research labs are anxious to share – or even make transparent – underlying data sources, is symbolic, and shows how nervous business actors are to share data. A position paper on the European Union’s Data Act published by Digital Europe trade association argues that B2G data sharing should be allowed only in the case of public emergencies.⁶

Public interest, while easily grasped intuitively in the context of data as the right of the public to know about things and obtain information relevant to this public, needs to be more easily conceptualized. In European digital policies, the term has not been adequately defined. In the debate on B2G data sharing, the term “general interest” was proposed instead, and was used to define a list of services that serve public obligations. Some examples of such services include healthcare or public transport. It should be noted that the public interest category is much broader and not limited to specific public benefits.

The difficulties with defining public interest in data sharing are one of the reasons that private actors are anxious to share data, seeing the data sharing provisions as a gateway to the data they hold being used in varied, unpredictable ways. The Data Impact Hub framework should therefore provide actors sharing data with more clarity on how the data will be used. A more precise public interest framework should be developed for the Data Impact Hub, since securing public interest

² Thomas Ramge and Viktor Mayer-Schoenberger, *Fuori i dati! Rompere i monopoli sulle informazioni per rilanciare il progresso* (Egea, 2021).

³ United Nations, “UN 2.0 Forward thinking culture and cutting-edge skills for better United Nations system impact,” *Our Common Agenda Policy Brief 11* (2023). Accessible at: <https://www.un.org/sites/un2.un.org/files/our-common-agenda-policy-brief-un-2.0-en.pdf>

⁴ Tiina Harkonen, Rita Vanshu, Jukka Vahti, and Kristo Lehtonen, *Tracking Digipower: How data can be used for influencing decision-makers and steering the world* (SITRA, 2022). Accessible at: <https://www.sitra.fi/app/uploads/2022/05/sitra-tracking-digipower-v3.pdf>.

⁵ Tim Davies and Silvana Fumega, *Global Data Barometer – First Edition* (2022). Accessible at: <https://doi.org/10.5281/zenodo.6488349>

⁶ Digital Europe, “Joint Statement: The Data Act is a leap into the unknown,” 1 February 2023, <https://www.digitaleurope.org/news/joint-statement-the-data-act-is-a-leap-into-the-unknown/>.

goals offers the most robust mandate for B2G data sharing. This framework should offer means of precisely conceptualizing public interest uses of data, but also offer risk assessment and accountability frameworks.

A commons-based perspective offers tools for securing public interest in data by offering a framework that moves away from the dominant, commercial models of data governance for private gain. The Public Data Commons aims to address challenges to public interest data sharing by proposing a systemic approach based on three key elements:

- Clearly defined public interest criteria for sharing data held by private bodies (and other data holders);
- Data governance mechanisms that include both Open Data and other data governance mechanisms based on gated access;
- A stewardship body that serves as a trusted party that ensures data availability and also acts as a gatekeeper that stewards the data commons.

As a result of applying these elements to the institutional, legal, and technical design of the Public Data Commons, three key goals can be achieved:

- The Public Data Commons can serve as a trusted intermediary for actors, both commercial and non-commercial, that makes data available for public interest sharing;
- The Public Data Commons plays a gatekeeping function that can accommodate various types of data and deploy multiple modes of governance, from Open Data sharing to gated access aimed at preserving individual and collective data rights;
- The Public Data Commons model allows safe and trustworthy storage of data, with data releases being triggered by the needs of combatting emergencies, yet with data accessible for reuse on a more systemic basis.

The Public Data Commons approach expands on the basic idea of B2G data sharing in emergency situations, where specific data-sharing requests are legitimized by a need to deal with a specific, current emergency. Firstly, it provides an institutional framework that is in line with the vision of pre-positioned capacities for data sharing. Furthermore, it aims to move beyond the logic of addressing specific, current emergencies through data sharing and instead provide a more systemic data-sharing approach. Data shared in this way would be securely stored or aggregated by the Public Data Commons and then made available not just for the purpose of dealing with the given emergency but also for future cases where public interest necessitates data use. The Public Data Commons institution would not only secure data rights but also make sure that the use is proportionate to alleviate the worries of commercial actors.

Open Future Foundation has published a *Data Commons Primer*,⁷ which offers a more fine-grained set of design principles for establishing commons-based data stewardship, and offers

⁷ Alek Tarkowski and Jan J. Zygmuntowski, "Data Commons Primer," 20 September 2022, <https://openfuture.eu/publication/data-commons-primer/>.

policy considerations for the Data Impact Hub. The aim of these principles is to balance public interest, generation of economic value, and care for fundamental rights. These are achieved through three main design principles: Stewarding Access, Collective Governance, and Public Value.

Lessons Learned from European Experiences with Data Governance

The Data Impact Hub will face the challenge of gradually promoting and mandating B2G data sharing to avoid the perception that through the hub, all commercial data will be mandated to be shared. One solution to this challenge is offered by the European Open Data policy framework, which includes the concept of high-value datasets.⁸ These are specific types of data, or even specific datasets that are shared as open data, based on regulatory decisions that take into consideration the public interest value of sharing different types of data. This type of focused, purpose-driven approach can mitigate fears around data sharing by limiting it to specific types and datasets - while at the same time ensuring that they have high potential to generate public value. As part of this mechanism, a cost-benefit analysis method has been developed and could serve as a reference point for similar data selection mechanisms by the Data Impact Hub.⁹

More generally, the public interest data-sharing frameworks considered in this brief build on the principles established by Open Data policies and on lessons learned through their implementation. At the same time, limitations of the Open Data model are clear when different types of data are considered – especially for datasets that include personal data or other types of sensitive data. Commons-based approaches to data governance, therefore, cover a broad spectrum of methods that offer more fine-grained governance mechanisms aimed at balancing public interest in data sharing with other considerations, related in particular to the protection of individual and collective data rights.

Some of these mechanisms are technical, such as privacy enhancing technologies (PETs). Others are legal, such as the public interest conditions for processing data, included in the European GDPR, the FAIR data principles¹⁰ adopted in the research sector, and the CARE data principles. Finally, there are also various institutional data-sharing frameworks being explored, such as data trusts, data spaces, or data unions. The Data Impact Hub should have the capacity to operate with the various modes of data governance, and function as a competence center capable of encouraging other actors to adopt them.

The Public Data Commons approach stresses the importance of an institution designated to steward the pool of data and serve as a gatekeeper managing access to the data. This is different

⁸ European Commission, “Commission defines high-value datasets to be made available for re-use,” 20 January 2023, <https://digital-strategy.ec.europa.eu/en/news/commission-defines-high-value-datasets-be-made-available-re-use>.

⁹ European Commission, *Identification of data themes for the extensions of public sector High-Value Datasets* (2023). Accessible at: <https://op.europa.eu/en/publication-detail/-/publication/3df0f526-2510-11ee-94cb-01aa75ed71a1/language-en>.

¹⁰ RFORCE11, The Fair Data Principles, last accessed 31 October 2023, <https://force11.org/info/the-fair-data-principles/>.

and more complex than the stewardship model developed for Open Data repositories. In the Public Data Commons model, an institutional actor is needed to serve as gatekeeper for two reasons. Firstly, requests to obtain access to data need to be reviewed so that public interest criteria are met. Secondly, gatekeeping functions are necessary so that data is shared in ways that protect the rights of data subjects. The standard data sharing platform is no longer an Open Data repository but rather a system based on such functionalities as gated APIs and federated learning mechanisms.¹¹

Most importantly, the data stewardship partner needs to function as a trusted institutional actor that will ensure that the interests and rights of data holders, data subjects, and data users are met. This trusted party status is necessary for the impact hub to successfully obtain or aggregate data, and then manage it and make it accessible. European public debates on B2G data sharing have shown that any such institution needs to guarantee protection of digital rights, privacy in particular, and to overcome some level of mistrust towards State institutions (and also public international bodies) when it comes to data sharing.¹² In the case of the UN Data Impact Hub, it will also need to overcome geopolitical challenges related to data sharing at the global level. A telling example of such tensions comes from the UK, where experts are questioning the current data sharing policies of the UK Biobank, arguing that access should be limited for researchers from China, over concerns surrounding the sharing of genetic data without reciprocity by Chinese institutions.¹³

Currently, the European Union is finalizing its legislative process on the Data Act, which will provide a cross-sectoral data governance framework. On top of this baseline framework, a set of sectoral governance frameworks is being developed, based on the concept of data spaces: federated data ecosystems within a certain application domain. Last year, a legislative proposal for a European Health Data Space was presented, including a robust mechanism for health data sharing. Governance mechanisms proposed for the European Data spaces can serve as further reference points for the Data Impact Hub.¹⁴

¹¹ Irene Solaiman, *The Gradient of Generative AI Release: Methods and Considerations* (2023). Accessible at : <https://doi.org/10.48550/arXiv.2302.04844>.

¹² Zuzanna Warso, "To succeed, data commons advocates must address privacy concerns head-on," Open Future, 2 March 2023, <https://openfuture.eu/blog/to-succeed-data-commons-advocates-must-address-privacy-concerns-head-on/>.

¹³ Shanti Das and Vincent Ni, "Fears over China's access to genetic data of UK citizens," The Guardian, 20 August 2022, <https://www.theguardian.com/science/2022/aug/20/fears-over-chinas-access-to-genetic-data-of-uk-citizens>.

¹⁴ L. Nagel and D. Lycklama, "Design Principles for Data Spaces," *Position Paper* (Berlin: Open DEI, 2021). Accessible at: <https://design-principles-for-data-spaces.org/>.

Key Considerations for the Data Impact Hub

From Emergency Response to Pre-positioned Data Sharing

The European Data Act limits B2G data sharing to situations of public emergency and exceptional need. Such a narrow scope might make such data sharing more feasible in light of the reluctance of business actors to do so. At the same time, provisions that are limited to situations of exceptional need cannot be seen as a systemic approach that mitigates the abovementioned power asymmetries in data. As proven by experiences from the Covid-19 pandemic, data sharing triggered by a state of emergency tends to come a moment too late.

For this reason, ad hoc data-sharing solutions in face of emergencies will not be sufficient. The Public Data Commons offers a model in which a trusted institution acting in the public interest not only pre-positions capacities but over time builds a body of data that can be shared securely and in a trustworthy manner. At the same time, a systemic approach needs to provide stronger guarantees to all parties that the data will be properly governed. The Tehdas project, which explored the feasibility of creating a European Health Data Space, has developed a sustainability model that can be applied beyond the field of health data. The model is based on five principles: legal basis for data use and robust governance, access to quality data, capacity and competence of data managers and users, funding, and trust.¹⁵

Incentivizing Business Entities to Share Data

Incentives to engage businesses in public interest data sharing of their privately-held data are a major issue that needs to be addressed while the Data Impact Hub is designed. On the one hand, there is a growing sense that data asymmetries can only be dealt with through State regulation. European Union rules allowing access to platform data for research purposes, included in the Digital Services Act, are an important expression of this idea. B2G data-sharing mandates are being explored in India¹⁶ and the UK, and have been recommended by the OECD in its report on Data Driven Innovation.¹⁷ On the other hand, calls for sharing private data face uniform criticism from market actors. In the face of this opposition, data-sharing strategies will need to take into account, in the long term, the issue of incentivizing market actors to share data.

In order to address this, the Data Impact Hub should first explore strategies for voluntary data sharing and deploy mechanisms that facilitate it. In Europe, a status of ‘recognized data altruism organization’ was introduced in the Data Governance Act for this purpose.¹⁸ Secondly, the above-

¹⁵ TEHDAS, Sustainability Plan for Secondary Use of Health Data in the European Health Data Space (2023). Accessible at: <https://tehdas.eu/app/uploads/2023/09/tehdas-sustainability-plan-for-ehds.pdf>.

¹⁶ Government of India, *Report by the Committee of Experts on Non-Personal Data Governance Framework* (2019). Accessible at: https://static.mygov.in/rest/s3fspublic/mygov_160922880751553221.pdf.

¹⁷ OECD, *Data-Driven Innovation: Big Data for Growth and Well-Being* (OECD, 2015). Available at: <https://www.oecd.org/sti/data-driven-innovation-9789264229358-en.htm>.

¹⁸ EU register of recognised data altruism organisations is accessible at: <https://digital-strategy.ec.europa.eu/en/policies/data-altruism-organisations>.

mentioned approach based on identifying specific high-value datasets or categories of data allows for gradual and focused data-sharing processes. Thirdly, mission-driven approaches allow the formulation of data-sharing strategies that highlight data sharing as a collaborative and purposeful effort.

Finally, it seems that consensus is most easily reached on research-related purposes for data sharing, as again proven by the European policy debate on the Digital Services Act. A strong research mandate should underpin the mutual commitment framework; while at the same time the hub should encourage data sharing for other public interest purposes as well.

Accounting for the Diversity of Data Sources and Types

Part of the challenge faced by proponents of B2G data sharing in the European policy debate was that the issue of sharing commercial data for public interest purposes was considered separately, as a dedicated B2G data-sharing mechanism was being negotiated. Considering B2G data sharing on its own is a difficult policy ask, since the business sector sees it largely as a regulatory burden and risk to competitiveness in a data-driven economy.

A more politically sound approach should consider the full spectrum of data sources, including publicly-generated data, sources of data that are open or voluntarily shared, and commercial data. A mission-driven approach, focused on defining clear goals for sharing data and means of collectively attaining them, offers a framework for sharing these various kinds of data by different actors, through a shared multistakeholder effort.

An example of such an approach is the Finnish Forestry Act, which establishes a public body tasked with gathering forestry data.¹⁹ The body has access to public data, has a B2G data-sharing mandate, and can also conduct data crowdsourcing efforts.

Another example is that of Findata, the Finnish Social and Health Data Permit authority, a public body that aggregates and makes available in a secure manner health data. While most of the data is sourced from different public sector sources, it also includes data from private social and health service providers.

Such multistakeholder data collection and data-sharing missions can be particularly successful if focused on addressing a specific gap in data availability, or a specific emergency that can be addressed by data-driven solutions. A broader data-sharing perspective will not solve all challenges faced by B2G data-sharing mandates, but they provide a political narrative that may facilitate the process.

¹⁹ See forest legislation in Finland: <https://mmm.fi/en/forests/legislation>.

Multilateral Governance of Data and State Actors as Trusted Parties

The European public debate on B2G data sharing demonstrated a high level of distrust towards State actors when it comes to stewarding data, especially when it is sensitive. Although there is growing agreement of a need to deal with data asymmetries caused by the concentration of data in the hands of business actors, the idea that broad swathes of data would be made available to and managed by governments remains controversial. In addition, the Data Impact Hub will face challenges related to multilateral arrangements, particularly since countries participating in multilateral processes do not always trust each other.

Measures should be adopted to ensure that the Hub is perceived by data owners and data subjects as a trusted party. This can be achieved by ensuring the protection of data rights (privacy in particular) and regulatory guardrails (for example, the Data Impact Hub should not be used for security purposes). Of additional importance are multistakeholder and participatory governance models with relation to the data aggregated and made available by the Hub.

Data Stewardship as an Institutional Function

Research into data sharing shows that institutional and technical solutions cannot be considered without dealing with the issue of individual and institutional skills and capacities. GovLab defines data stewards as “organizational leaders or teams empowered to create public value by re-using their organization’s data.”²⁰ The Data Impact Hub should be seen as a data stewardship institution functioning within the global, multilateral system. Data-sharing policies should be anchored to dedicated institutions that offer legal and regulatory expertise and provide necessary data-sharing infrastructures. They also play the role of an institutional leader for other organizations, guiding them through a process of institutional change that will make them more capable of properly managing and sharing data.

Beyond Data Governance and Towards Public AI Capacities

In the last year, a growing number of data governance debates have occurred in conversations about the regulation of new machine learning (artificial intelligence) technologies. While machine learning is not the sole form of data use that should be considered, it constitutes an emergent form that is relevant not just because of its potential impact, but also because of specific risks and challenges to data governance that are being identified. It can be expected that in the coming years an increasing number of data-driven efforts will be based on machine learning technologies.

²⁰ GovLab, *Wanted: Data Stewards: (Re)Defining the Roles and Responsibilities of Data Stewards for an Age of Data Collaboration* (2020). Accessible at: <https://thegovlab.org/static/files/publications/wanted-data-stewards.pdf>.

For this reason, a future-proof design of the Data Impact Hub should also take into consideration issues related to the governance of machine learning tools and other algorithmic services that either could be deployed as part of the Hub or that will constitute a typical means of accessing and using the Hub's data. While the governance challenges and solutions will not be different, it is worth exploring whether the Data Impact Hub model should also include a public option for AI: democratically-governed and publicly-accessible models that would serve the same purpose as the data sharing mechanisms.

United Nations University - Centre for Policy Research

767 3rd Ave Floor 35, New York, NY 10017

This Working Paper is an output of the UNU-CPR initiative, “A Breakthrough for People and Planet: Building Momentum for the Summit of the Future and Beyond”.

