COMMUNITY OF PRACTICE REPORT

Exploring Data as and in Service of the Public Good

Co-authored by: the Digital Public Goods Alliance, Global Partnership for Sustainable Development Data, Jain Family Institute, UN Global Pulse, and UNICEF

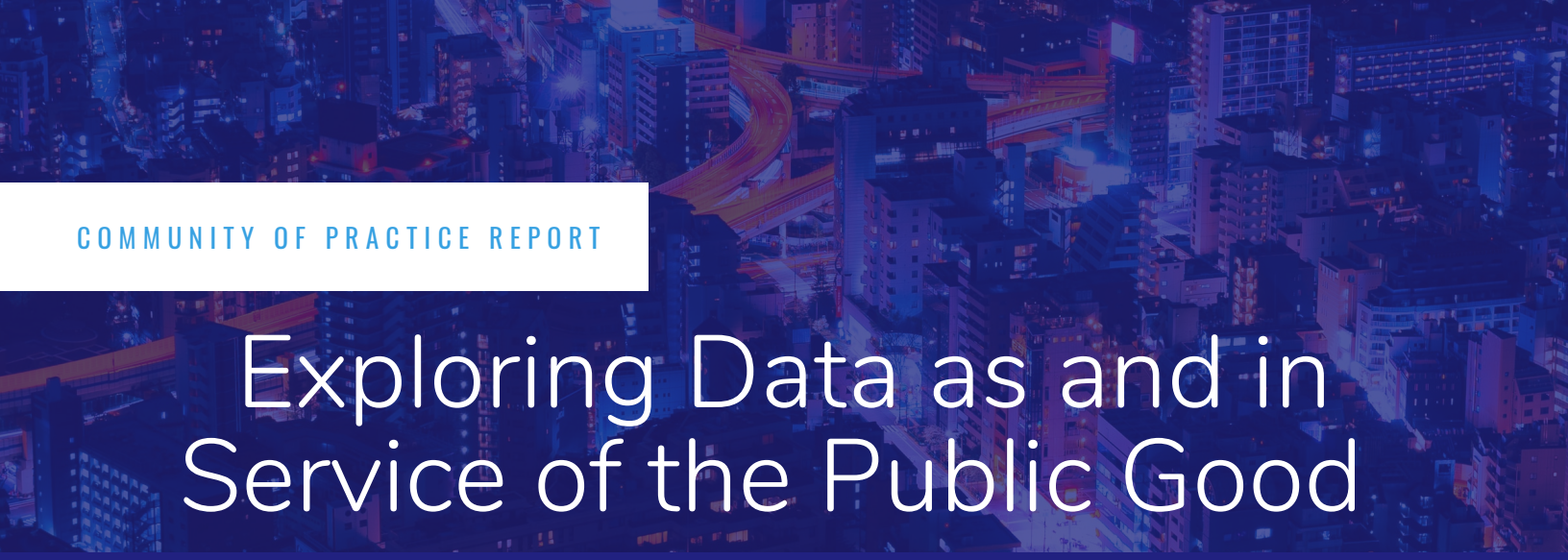
Digital Public Goods Alliance

Global Partnership for Sustainable Development Data

Jain Family Institute

UN GLOBAL PULSE

unicef for every child



COMMUNITY OF PRACTICE REPORT

Exploring Data as and in Service of the Public Good

Established in 2019, the Digital Public Goods Alliance is a multi-stakeholder initiative with a mission to accelerate the attainment of the sustainable development goals in low- and middle-income countries by facilitating the discovery, development, use of, and investment in digital public goods. Digital public goods are open-source software, open data, open AI models, open standards, and open content that adhere to privacy and other applicable laws and best practices, do no harm by design, and help attain the SDGs. To learn more, visit digitalpublicgoods.net or contact hello@digitalpublicgoods.net

Cover photo credit: Pawel Nolbert

COMMUNITY OF PRACTICE REPORT

Exploring Data as and in Service of the Public Good

Table of Contents

Background

The goal of this paper is to explore when data can be made openly available as a digital public good (DPG) and propose how crucial data governance questions might be approached in situations where data cannot be made open but still serves the public good. The analytical framework presented here is intended to serve as a foundation for future discussion. The questions raised act as a roadmap to move from diagnosis to solution, with the goal to unlock the value of data in service of the public good.

This paper was Co-Authored by Digital Public Goods Alliance (DPGA), Global Partnership for Sustainable Development Data, Jain Family Institute, UN Global Pulse, and UNICEF as part of a community of practice on "Data for the Public Good," in coordination with the UN High Level Committee on Programmes (HLCP) working group on new global public goods: international data governance, which is led by the Committee of Chief Statisticians of the UN (CCS-UN).

Communities of practice bring together experts from different institutions and networks, who meet to discuss how digital public goods might be better deployed to address critical development needs and challenges. The work of this community of practice provides the foundation for a submission to the UN's Global Digital Compact, and may result in a revision of the DPG Standard moving closer towards realizing the promise of data as a DPG.[1] Many thanks to the community of practice members listed below for their contributions:

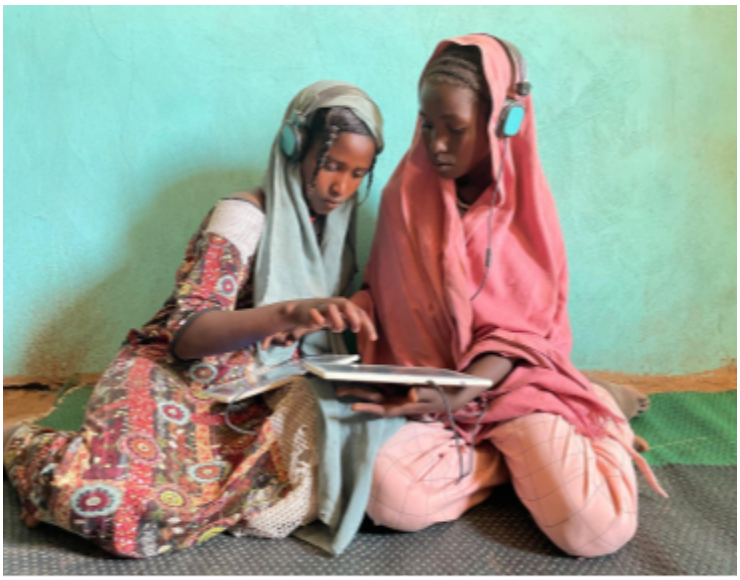
- Ahmed El Saeed, UN Global Pulse
- Alex Diaz, Google
- Angela Me, UNODC
- Arden Ali, The Jain Family Institute
- Caroline Alewaerts, UN Global Pulse
- Claudia Juech, Patrick J. McGovern Foundation
- David Jensen, UNEP
- David Passarelli, UNU-CPR
- Deepa Karthykeyan, Athena Infonomics
- Denise Mckenzie, PLACE
- Em Lewis, Mozilla
- Emanuel Moss, Intel Labs
- Eleonore Fournier-Tombs, UNU-CPR
- Friederike Schüür, UNICEF
- Gianfranco Cecconi, Capgemini Invent
- Holly Krambeck, The World Bank
- Jameson Voisin, Digital Public Goods Alliance
- Jay Hodges, Tean Jain Family Institute
- Jean-Martin Bauer, WFP
- Jenna Slotin, Global Partnership for Sustainable Development Data
- Jennifer Pratt Miles, The Lacuna Fund
- JoAnn Stonier, Mastercard
- Lea Gimpel, Digital Public Goods Alliance
- Lucy Harris, Digital Public Goods Alliance
- Mala Kumar, GitHub
- Maria Rosaria Conduti, European Commission
- Mark Hereward, UNICEF
- Martina Barbero, Global Partnership for Sustainable Development Data
- Maximilian Gahntz, Mozilla Foundation
- Minke Meijnders, UN Global Pulse
- Natalia Carfi, Open Data Charter
- Peter Rabley, PLACE
- Philipp Schulte, Deutsche Gesellschaft für Internationale Zusammenarbeit GmbH (GIZ)
- Rohan Samarajiva, LIRNasia
- Siddharth Shetty, iSPIRT
- Sonia Cooper, Microsoft
- Sriganesh Lokanathan, LIRNasia
- Stefaan Verhulst, NYU
- Stephen Mac Feely, WHO
- Tiina Neuvonen, UN Global Pulse
- Vivek Sakhrani, Atlas AI
- Yolanda Lannquist, The Future Society

[1] Digital public goods refer to open-source software, open AI models, open standards, open content, and open data that adhere to privacy and other applicable international and domestic laws, standards, best practices, and do no harm. The concept of DPGs stems from the economic term "public good" referring to resources and services individuals cannot (or should not) be excluded from. The DPGA stewards this definition. The DPG Standard is a set of specifications and guidelines designed to maximize consensus and establish baseline requirements that must be met for digital solutions to be considered DPGs.

Executive Summary

Data is becoming a key driver of economic vitality with demonstrated potential to serve the public good. It can play a key role in the fulfillment of human rights, including child rights, and attainment of the Sustainable Development Goals (SDGs). However, the lack of timely and reliable access to quality data is a barrier towards realizing its potential in this regard. Therefore, we have a collective obligation to take action. One option for removing impediments is to apply an open-source license[2] to a dataset. This will make it accessible to all with very minimal restrictions. Indeed, data can be considered a digital public good when there is open-source licensing combined with the appropriate documentation and alignment with the DPG Standard.[3] However, there are many datasets which cannot be shared with the public because full open access would be at odds with public benefit, for example when there are security risks. In these situations, the data may be best managed akin to a "club good."

The club good structure resolves trade-offs between (a) data access and protection; (b) promotion of data use for commercial and public benefit; and (c) fair distribution of risks and benefits derived from data, across individuals and communities. However, data club membership rules need to be carefully designed to allow appropriate actors to (a) realize public benefits through data use; (b) minimize accidental misuse; (c) not engage in intentional misuse; (d) participate in decisions on who, how, and when data can be used; (e) adhere to the public good use of data; and (f) collectively decide how to fairly distribute public benefits across individuals and communities. With carefully designed membership and decision rules, data clubs could emerge as a suitable structure for governance of data as a resource, as well as serve as a data commons in service of the public good for data where an open source license approach is not appropriate.



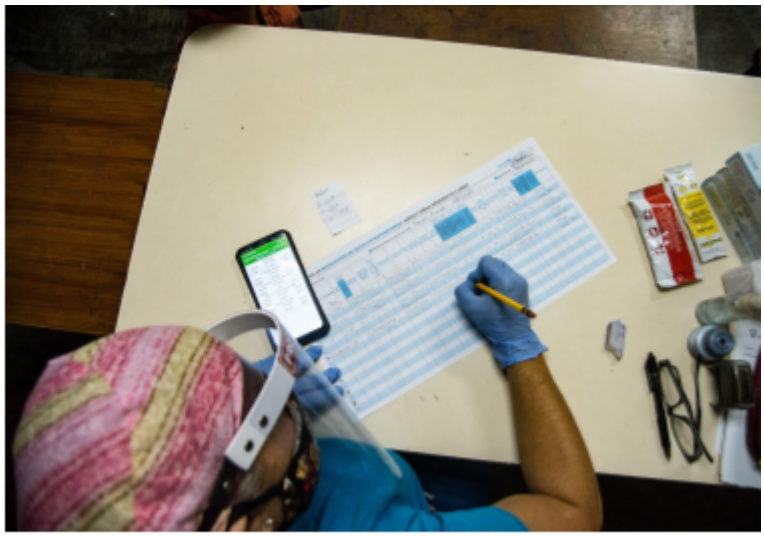
© UNICEF/UN0753327/Bos

[2] Open Definition, "Creative Commons Attribution License," Open Knowledge Foundation, https://opendefinition.org/licenses/cc-by/

[3] "DPG Standard," Digital Public Goods Alliance, https://digitalpublicgoods.net/standard/

Introduction

Data is playing an increasing role in the public and private sector. In the private sector, data serves to improve the quality and reduce the cost of delivery of products and services, while fostering innovation, including the development of new products and services. Data is becoming a key driver of economic vitality and well-being. In the public sector, data improves policy making and service delivery by helping to channel scarce resources to those most in need, including children (particularly during humanitarian crises); providing the means to hold governments accountable; and fostering social innovation. In short, data has the potential to improve people's lives.[4]



© UNICEF/UN0374178/Cabral

However, there are challenges when it comes to responsibly harnessing the value of data. In the commercial sphere, it remains unclear what kinds of data-use constitute fair competition at national, regional, and global levels. There are also questions about who should–and should not—benefit from data. When it comes to personal data, it is also important to consider whether the benefits derived appropriately reach the individuals and communities that the data is about and/or collected from.

Presently, the main impediment to harnessing the full potential of data for the public good is the lack of timely and reliable access to quality data. In part, this is due to a lack of investment in the collection and quality control of public-intent data, as well as a lack of investment in capacity building to make effective use of available data. Further, public access to private-intent data is often lacking.[5] There are often restrictions and legal uncertainty on the reuse of data. Finally, it must be acknowledged that data has the potential to cause harm: it can erode privacy, be used for manipulation and undermine individual autonomy.

In order to move the conversation forward, we need to think about how data should be governed. Two key approaches are: (a) positioning data as a digital public good; or (b) governing data like a club good. In this paper, we conclude that public-intent data should be made as a digital public good in those cases where it can safely be made openly available. However, in those cases where data requires more fine-grained governance, it should be managed as a club good, but with particular considerations on club membership, rules, and mandate that preserve a focus on the public good.

To facilitate the discussion, we introduce a new concept:

> Community data is any dataset that has the potential to play an important role in the fulfillment of human rights and attainment of the SDGs, that individuals and communities have a right to benefit from, and that needs to be accessible to relevant actors in ways that minimize the risks of accidental or intentional infringement of human rights or otherwise set back progress towards the SDGs.

[4] *Data for Better Lives* by the World Bank (2021) provides deep coverage of the commercial and public value of data and details how to overcome barriers towards changing the data landscape to improve the lives of poor people: https://www.worldbank.org/en/publication/wdr2021

[5] i.e. data collected and curated as part of business routine processes by the private sector for commercial purposes

Laying out the conceptual framework for digital public goods and club goods

How data should be governed can be best determined by its characteristics as a resource. Resources are broadly assigned two fundamental characteristics: they are either "rivalrous" or "non-rivalrous" and either "excludable" or "non-excludable."

A resource is rivalrous if use by one consumer prevents use by other consumers. Clearly, this is not the case for data. The very same data set can be used by multiple consumers without its value diminishing. Data sets can even be used simultaneously by multiple people for different purposes. Furthermore, especially when data is digitized, it is possible to set up systems so that the monetary cost of sharing it among additional users is (near) zero. For these reasons, data is a non-rivalrous good.

Excludability is defined as the degree to which access to the resource can be limited or, conversely, the degree to which a producer or other managing body can prevent its "free" consumption. Given that data is often withheld from public access (and therefore public use), it is currently treated mostly as an excludable resource. This is particularly true for data held by private sector entities, who limit access and use of their data.

When a good is non-rivalrous but excludable it is considered a "club good." Consequently, data that is not made openly available is treated as a "club good" with club membership currently largely determined by who collects the data. For private-intent data, members of the club are largely corporate entities. They have access to the data and therefore derive its benefits.

	EXCLUDABLE	NON-EXCLUDABLE
RIVALROUS	Private goods	Common pool goods
NON-RIVALROUS	Club goods	Public goods (Digital public goods)

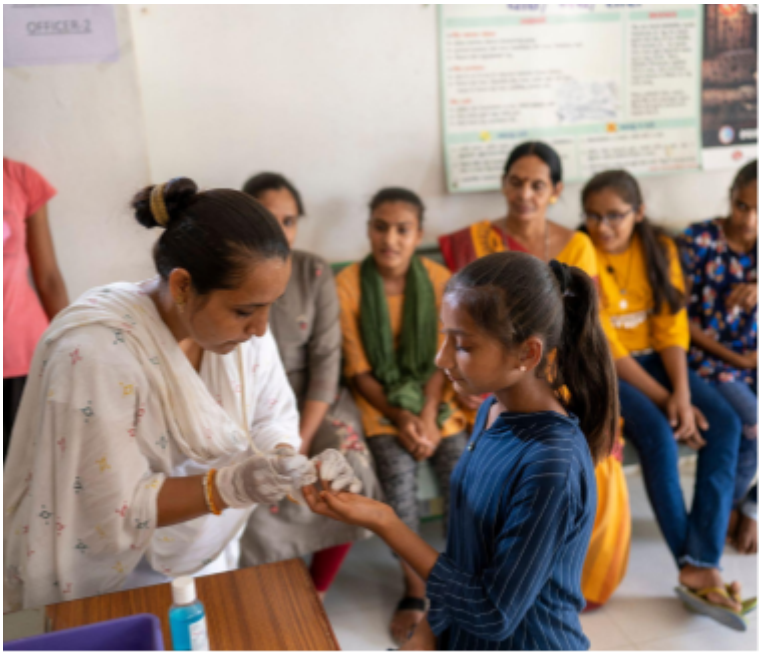
Excludability is determined by political will and technological innovation. There are efforts and mechanisms to make data more non-excludable, for example, through the use of open-source licenses, which facilitate free and unrestricted access. In these cases, the data can be considered as a digital public good since it is both non-rivalrous and non-excludable. However, a vast majority of data, both public-intent and, more predominantly, private-intent data, is not available under an open-source license— this includes key community data. Therefore as we continue to move relevant data sets towards openness and digital public goods, in order to fulfill our collective obligation to realise the potential of data for public good it is also critical to explore other governance models. The club good concept opens the possibility to interrogate excludability through club membership and propose new expectations around who is involved in deciding whether, how, and with whom data is shared, and who benefits from data use.

Current efforts and attempts at solutions

Data as a digital public good

There is a collective obligation to both provide access to and protect data. Managing data as a fully non-excludable resource with no limitations on access, such as through an open-source license, removes barriers as well as maximizes use and reuse.

When a data set is openly accessible, a potentially large number of people can contribute to it and thereby increase its quality, which in turn increases the likelihood that data can play a key role in the fulfillment of human rights and the attainment of the SDGs. For example, the Norwegian Meteorological Institute provides high-quality, reliable, and easy to use ten-day weather forecasts for any geolocation in the world under an open-source license. As a result, communities worldwide can learn from the data and make vital weather-related decisions which can help improve lives.



© UNICEF/UN0652789/Panjwani

Managing data as a fully non-excludable resource, or as a digital public good, makes it accessible for reuse, thereby likely increasing the availability of key community data sets, paving the path towards realizing the promise of community data. However, management of data as a fully non-excludable resource, or digital public goods, opens up possibilities for its accidental or intentional misuse, which may erode human rights and set back attainment of the SDGs. For example, the high-resolution satellite imagery data managed by PLACE is designed to deliver timely, quality, and consistent mapping data for the public good.[6] However, because sharing detailed data on sensitive locations could pose security risks, they help manage controlled access.

The challenge, then, is how to ensure that data is made as openly accessible as possible while still limiting access to data with the potential to do harm. Some data sets without the potential for accidental or intentional misuse could be managed as a digital public good. However, data sets that contain sensitive information or otherwise carry potential for misuse need to be managed. The concept of a club good is useful. Access to these data sets would be restricted to members, who would be obligated to use it to serve the public good.

Given their privileged role in accessing and using this restricted data, it is important to select club members carefully. We believe that the conditions for club membership must be such that actors (a) can realize public benefits through the use of data, including those who the data is about; (b) can minimize accidental misuse of data; and (c) do not engage in intentional misuse; (d) can gain membership to the club without excessive efforts. To answer these needs, PLACE developed a membership model that controls access to data sets. To gain access to these sets, members

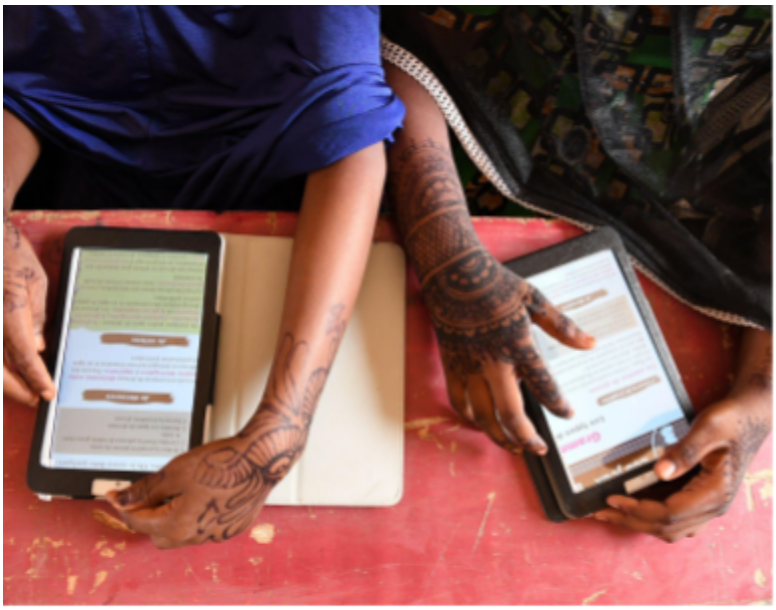
[6] "Introducing PLACE: mapping data in the public interest," PLACE, https://www.thisisplace.org/blog-1/introducingplace/introducing-place

must agree to ethical terms and conditions and they must also remain in good standing to maintain their access. Crucially, the benefits of data realized through the work of the members of the club need to be accessible to all as a public good which also helps the aspiration of community data.

Club goods for community data

To realize the aspiration expressed in the definition of community data, there are three key tensions to navigate: (a) the collective obligation to provide access and protect data for public benefit;[7] (b) the opportunity to use data for commercial benefit; and (c) the fair distribution of benefits derived from data across individuals and communities, and relatedly, the risks inherent in the use of data imposed on the individuals or communities. Determining the appropriate trade-offs needs to be part of any adequate mechanism for community data.

Data stewardship has become a popular way of thinking about how to manage these trade-offs. While public and private entities have different ways of describing data stewardship, it broadly refers to a function or set of functions to facilitate the production, management, sharing, and use of data within and between organizations in a responsible and trustworthy manner.[8] Managing data as a club good is one form of data stewardship. Recent experimentation with new data governance mechanisms, including data trusts, data cooperatives, and data collaboratives offer insights regarding how the club rules could be established and what approaches may be most effective for different types of data.



© UNICEF/UN0318708/Frank Dejongh

Governing club goods

One promising solution to realize the aspiration of community data is managing access and use of data as a club good while ensuring that benefits derived from data are managed as a non-excludable public good. With benefits derived from data as a public good, no individual or community is excluded from enjoying these benefits. However, attention needs to be paid to who gains club membership. Unequal barriers to club entry and participation in club decision making may influence what benefits are derived from data which, in turn, may lead to some individuals and communities benefiting from the data over others. A key consideration is the role that data subjects and affected communities play in the club, and how decisions are made about whether and how data is used. New innovations in data governance such as data trusts, data cooperatives, and data collaboratives offer different approaches for setting decision-making rules and ensuring that affected people and communities play a part in that process.

The club structure also allows enforcement of risk assessment tools, which identify risks and ensure proportionality of risks to benefits and maintain an acceptable distribution of risks across individuals and communities. As there are risks associated with opening up the use of public and private-intent data through club structures to realize the aspiration

[7] The obligation is to protect the "people and communities behind the data", to be precise, not the data itself.
[8] "Reimagining Data and Power: A roadmap for putting values at the heart of data," Global Partnership for Sustainable Development Data, https://www.data4sdgs.org/sites/default/files/2022-07/Final%20White%20Paper%20designed%20%28English%29.pdf

of community data. To start, data may inadequately represent populations (in fact, data sets may leave out certain populations altogether). This, in turn, may result in benefits that do not serve everyone equally, a risk heightened for private-intent data since much data collection of private-intent data is not designed to be representative of the population.

While these challenges exist regardless of the mechanism used to manage data as a resource, certain mechanisms may heighten these risks. For instance, managing data as a resource through markets is likely to result in a non-optimal distribution of risks and benefits across individuals and communities. This is because individuals facing greater economic pressures may end up overexposed to risks of data use.

Similarly, lack of representation (or bias) in data is likely to also arise in case of data philanthropy, especially if it is driven by corporate decisions to make available data gained through online participation and transactions with individuals. The club structure allows for tools to be put in place to identify lack of representation thereby mitigating risks of unequal benefits, unequal exposure to risks, and lack of proportionality of risks to benefits.

Outstanding questions and potential solutions for the governance of club goods

The solution that emerges for data governance to fulfill our collective obligation is that where open access poses no risk, data needs to be managed as a digital public good—with an open license as well as documentation and adherence to the DPG Standards with regard to access and reuse. Where there are risks, community data needs to be managed as a club good with carefully curated access. However, the public benefits derived from the use of community data within the club good structure needs to be managed as a (non-excludable) public good.

The section that follows raises crucial questions and offers additional considerations for them. They were gathered throughout the roundtables that produced this paper. Although more discussion is needed before the implementation of data clubs, the following should be considered:

1. What are the membership rules for public benefit data clubs?[9] Membership rules need to be designed such that the club provides data access and enables data use to diverse actors who (a) realize public benefits through data use; (b) minimize accidental misuse; (c) do not engage in intentional misuse; (d) adhere to the public good use of data; (e) collectively create fairly distributed public benefits across individuals and communities.

2. What are the decision-making rules for public benefit data clubs? Decision rules need to be designed to facilitate inclusive and participatory decision-making that ensure rights are respected, including consultations with data subjects, communities, and intended beneficiaries.

3. What are the enforcement mechanisms to prevent or respond to the infringement of membership rules? Enforcement mechanisms need to consider the use of data that causes harm or use of data for commercial benefits in violation of purpose specifications.

[9] The community-based approach to high-value datasets proposed by the Indian Government in its 2020 Report by the Committee of Experts on Non-personal Data Governance Framework (revised) may here serve as an example, with the proposed establishment of a data trustee to manage access and sharing of high-value datasets and tasked with the "exercise of the rights of the community over non-personal data collected in these high-value datasets".

4. Is there a single public benefit data club, or several? And at what level should these clubs be managed? Multiple approaches should be considered, in that data clubs could be created for specific sectors and/or at different levels (i.e., local, national, regional, or global).

5. How do we preserve and expand incentives for data collection and storage? There is a need to ensure solutions don't accidentally erode incentives through access to community data in data clubs.

6. What purposes for data use are within the scope of community data? There is a need to specify the precise role of data in the fulfillment of human rights and attainment of the SDGs, while also addressing the process of modifying public good purposes for data use over time, with the understanding that innovation may create new opportunities.

7. What data sets are community data sets? Who is involved in defining this? And how will we identify which should be digital public goods and which should be club goods?[10] There is a need to further define what it means for a data set to have the potential to play a key role in the fulfillment of human rights and attainment of the SDGs. For digital public goods, there is already a nomination process via the Digital Public Goods Alliance, which allows data owners to nominate their data sets as relevant to the SDGs. However, there is no similar process for identifying and supporting club goods. Grounding in human rights and the SDGs offers a starting point for addressing this critical question. Progress should be incremental, starting with datasets where the potential benefit for public good has been firmly established.

As part of the development of this report, a consultative foresight exercise was conducted resulting in four speculative scenarios describing the Future of Data Governance in 2050. The scenarios could serve as a tool for structural multi-stakeholder consultations and dialogues to contextualize the questions and potential solutions and strengthen long-term thinking around these issues.[11]

Conclusion

To create the data clubs and/or digital public goods that will enable us to realize the aspiration of community data in service of the public good requires political will, especially in light of the status quo and the entities that today have *de facto* control over the majority of data access and use. Despite the long and likely bumpy road ahead, openly licensed data sets and public benefit data clubs promise a path towards realizing the public good potential of data and towards living up to our collective obligation to fulfill human rights and attain the SDGs. This would help return the benefits of data back to individuals and communities.

[10] In this respect, inspiration could be found in the various approaches developed around the globe to identifying high-value datasets, such as the European Union approach (see the "High-Value Datasets" framework set out in the Directive (EU) 2019/1024 on open data and the re-use of public sector information (recast), in particular article 14, and its draft Implementing Act), the Canadian approach (see the "High-Value Datasets criteria" set out in the 2018 Report of the Canada Open Government Working Group), the Australian approach (see the proposed concept of "National Interest Datasets" under the Productivity Commission Recommendations 7.1 and 7.2 and the Australia's Open Government Partnership National Action Plan 2016-2018), or the the Indian approach (see the concepts of "Community" and "High-Value Datasets" under the 2020 Report by the Committee of Experts on Non-personal Data Governance Framework (revised)). It is however of note that, from these illustrations, only the Indian approach includes both data from the public sector and the private sector within the concept of high-value datasets, while the European, Canadian, and Australian approaches are narrowed to public sector datasets.

[11] The scenario report on the Future of Data Governance, and an accompanying scenario simulation game, will be published on the website of UN Global Pulse in Spring 2023.


Digital Public Goods Alliance
Global Partnership for Sustainable Development Data
Jain Family Institute
UN GLOBAL PULSE
unicef
for every child