Executive Summary

This case study was developed as part of *Effective Data Sharing: Beyond Platforms*, a study conducted by Athena Infonomics and commissioned by the Global Partnership for Sustainable Development Data. This project was made possible by a grant from Google.org. Special thanks to Vivek Sakhrani from Atlas AI who served as a senior technical advisor on this study.

Recent technological advancements have exponentially increased the availability of data, improved the efficiency of data sharing, and increased the ability to process and draw analytical insights from such data. The past two decades have been marked by a data revolution in the form of an ability to collect and analyse data. These advances present an opportunity to aid economic and social growth, both in the private and public sector. This data revolution presents immense potential for solving global development issues. Governments, international organisations, and other development actors rely heavily on data to plan and execute solutions to socio-economic problems.[1]

1 World Bank, Data-Driven Development, 2018

At the same time, digital advances present the risk of exacerbating existing social, economic, and geographic inequalities. Statistics suggest that less than 100 companies amass more than 50% of data collected online,[2] with these companies being concentrated in just a few countries. If this inequality is not addressed, the societal and economical value of improved data and computing will be reaped by a few economies and companies alone.[3]

2 Yokoyama, Closing the data divide: the need for open data, 2020

3 Microsoft, Sharing Data for Impact: Lessons From Data Sharing Initiatives in Asia, 2021

Definition of data-sharing partnerships

For this study, data-sharing partnerships are defined as partnerships where two or more organisations collaborate on data collection/ data sharing/data analysis to address a societal challenge. This study focuses on initiatives with three or more partners only, i.e., multi-stakeholder partnerships.

Data-sharing partnerships present a way to overcome gaps in data access and data capacities. These partnerships, when constructed safely, can provide a secure path for sharing data to achieve pre-decided objectives. Data sharing can serve societal good, but it can also help businesses identify new models that will promote economic and social benefits.[4] A study by the Open Data Institute (ODI) demonstrates that data sharing in the private sector **can encourage innovation, optimise supply chains, address sectoral challenges, improve market reach, help comply with regulation.**[5]

4 Russo, et al., Sharing Data to Address Our Biggest Societal Challenges, 2021

5 ODI, Data Sharing in the Private Sector, 2020

In the public sector, improved data-sharing **can improve public services, facilitate research and innovation, and inform policymaking.**[6] A 2013 study in the UK estimated the direct and indirect impact of public-sector data sharing to be around GBP 5 billion (USD 6.5 billion) per year. This estimate included direct revenues and indirect impact on data users such as time saved from access to real-time travel data.[7] The OECD estimates that improved data access and sharing could generate benefits valued at between 1% and 2.5% of the GDP.[8]

The COVID-19 pandemic has highlighted the promise, opportunities, and challenges these data-sharing partnerships hold. The pandemic pushed governments and private-sec-

6 Drummond and Christie, Sharing Public Sector Data, 2022

7 OECD, Enhancing Access to and Sharing of Data: Reconciling Risks and Benefits for Data Re-use across Societies

8 OECD, Enhancing Access to and Sharing of Data: Reconciling Risks and Benefits for Data Re-use across Societies

tor actors to work together to plug data gaps as traditional data-collection systems faced significant disruptions in their operations.[9] Governments worldwide relied on alternate data sources (such as mobile data records) and advanced analytics to inform decisions around mobility, health policy, and other areas.[10] Governments used big data models to track migrating populations, identify vulnerable groups, monitor lockdown compliance, and manage and optimize the usage of scarce resources.[11] Data sharing between public and private actors allowed access to pertinent data and informed crucial policies. This experience highlighted the need for incentivising and enabling structured data sharing.[12]

Despite the expected benefits of improved data sharing, concerns around data security and privacy, a lack of control of data shared, intellectual property rights, and commercial interests hinder data sharing. Concerns around intellectual property rights and commercial use of data are particularly important when accessing data from private entities. While there has been a proliferation of data-sharing initiatives in the past few years, many of these initiatives have had short lifespans and often become defunct in a couple of years.

This suggests that there is a gap in understanding the key enablers of data sharing, and there is a need for stakeholders to clearly evaluate the elements that contribute to successful data-sharing partnerships.

Image by rawixel.com on Freepik

9 UN Stats, How COVID-19 is changing the world: a statistical perspective, 2020

10 Hayden Dahmm, Data Sharing in a Post-Pandemic World: How to Safely Wind Down Surveillance Measures, 2020

11 GSMA, Utilising mobile big data and AI to benefit society, 2021

12 Jay Park et. Al., How COVID-19 has fundamentally changed clinical research in global health, 2021

About this study

While several data-sharing partnerships are emerging in the private and public sector, there have been limited efforts to systematically analyse them and identify lessons learned especially in the public sector. There has been an increase in the number of data-sharing partnerships formed, but a significant number of them dissolve quickly or fail to take off. It is clear that there is a need to better understand how to create effective and sustainable data-sharing partnerships.

This study investigated the landscape of data-sharing partnerships, identifying the key drivers of success. To achieve this objective, the study team developed a framework to systematically assess the key elements of establishing, operating, and maintaining a data-sharing partnership. Both secondary and primary research methods were used to develop, refine, and inform this framework. The framework incorporates learnings from resources and guides on data-sharing, including: the ODI's extensive work on data sharing; the EU Support Centre for Data Sharing; NYU GovLab's work on Data Collaboratives; The Data Economy Lab's work on data stewardship and governance; CABI's Data Sharing Toolkit; and the Mozilla Foundation's Data for Empowerment research.

Based on secondary research, a broad analytical framework based on three pillars – governance, sustainability, and data architecture – was developed. Specific dimensions were developed under each of these pillars to unpack its constituent elements. This framework was further refined through discussions with the reference working group, with the Global Partnership for Sustainable Development Data (GPSDD) team and with key informants. Questions for both primary and secondary research were mapped to each constituent dimension. This guided both the desk research and the key informant interviews.

The key learnings from the study are presented below

1 Regulatory compliance is a pre-requisite for establishing an initiative.

Compliance with regulatory frameworks is a necessary pre-condition for the establishment of any data-sharing initiative. Across the globe, regulators are catching up to technological advances in data sharing and the regulatory landscape is fast evolving. Data-sharing partnerships must consistently review relevant laws and compliance with them.

Design decisions on the type of data that can be used, the partners that can be brought into the initiative, and the location of the initiative are determined by applicable laws. **The type of data used by an initiative impacts its regulatory burden.** Data sharing involving personally identifiable information is regulated stringently across countries. However, once data is anonymised and aggregated, the data can be shared more freely. **Initiatives housed in international organisations, such as the UN, may find some exemptions from local and national regulations,** depending on the type of data and its use.

Regulations can play an enabling or disabling role in the success of data-sharing initiatives. Since most governments are still evolving their approach to data privacy and security, **a lack of clarity in regulations can be a cause of concern.** GDPR and the ASEAN data-management framework provide examples of regulatory co-operation that promote cross-border data sharing, and provide explicit rules on the boundaries of data sharing. At the same time, restrictive or overly detailed frameworks can make it difficult to access and use data.

2 Trust emerged as a minimum and necessary pre-condition for all data-sharing partnerships.

Across all data partnerships explored in this study, **establishing trust between data partners, and building trust in the operations of their respective initiatives emerged as the single most important driver of success.**

Trust building is a constant and iterative process, which remains a focus throughout the lifecycle of the data-sharing partnership.

While data sharing creates value, it can also expose the data sharers to harm by putting them at legal, reputational, or data risk. Data sharing and open data also pose a risk when accessed by 'bad' actors and/or those without a technical grasp of the subject matter. There is a **higher trust barrier when personally or financially sensitive data is being used. Initiatives that seek to inform government action or that use public data need to cultivate a higher level of trust from the government and citizens.**

Bringing the right partners to the table and getting them to share the data **can be a long, resource-intensive process.** Trust-building workshops, prior history of collaboration, an extended communication phase, and buy-in from organisational and sectoral champions can help establish trust amongst partners. **Initiatives housed in international organisations, such as the UN, work with an existing reservoir of trust which makes it easier to establish credibility and build data partnerships.**

3 Internal governance norms can help build transparency and accountability.

Organisational and data governance norms can aid the achievement of an initiative's objectives, build transparency and accountability within its functioning, allow stakeholders to participate in decision-making, and provide clarity on data management. **At its core, these processes seek to build continued trust and collaboration between partners.**

Internal governance mechanisms try to strike a balance between control and trust. The more trust there is between partners, slightly lower control might be exercised by the internal governance arrangement. However, when there is limited trust between the partners, more rigid organisational and data governance rules will be in place, exercising greater control on the functionining of the initiative. **Greater control is sought if sensitive data is being shared.** Flexibility in internal governance can lead to growth and innovation.

4 Data science knowledge is often valued more highly than sectoral expertise or an understanding of the socio-political climate in which the data model is operating.

Investments in building the capacity of internal and external stakeholders is important for accountability, sustainability, and usability. This can occur within an organisation (identifying staff with the right capacities); with partners (investing in improving the data capacity of partners or communities that the initiative works with); or through general knowledge management (creating data-sharing resources for research, academic, and data communities).

Most data-sharing initiatives interviewed for this study found that data science capacity needed for advanced data-sharing and processing was concentrated within certain populations and geographies. Across interviews conducted for this study, **there was an acknowledged need to build data science skills across diverse populations, especially in the geographies that these initiatives were trying to impact.**

Across different case studies, it was clear that data science skills were valued more highly than local knowledge and understanding. While the former skills tended to be hired directly into the initiative, gaps in local context and knowledge were overcome by measures such as hiring national consultants. However, there were **concerns expressed that the lack of adequate staff in the implementing geography adversely impacted operations of the initiative and the data model.** Communicating the value of different capacities is important in attracting investments, time and money – into developing varied skillsets within a data-sharing partnership.

5 Availability of long-term, stable funding is crucial for the growth, innovation, and sustainability of data-sharing partnerships.

The ability of an initiative to generate funds or revenue impacts the sustainability of its operations and investments in continued growth. Ensuring financial sustainability and exploring a diverse set of revenue models was a common concern across most data-sharing initiatives studied. **While a majority of the initiatives remain donor-funded, all of**

them recognised the importance of building revenue models into their operations. While most initiatives expressed interest in moving from donor-based to revenue-generating (or partially revenue-generating) models, there were limited examples of successful transitions.

Models that were being explored included subscription fees; fees for data services offered; fees for data storage or processing, etc. However, such models are still in their early stages and their scale and sustainability is unclear. A limited number of initiatives were able to raise revenue to sustain their ongoing operations. However, donor funding tended to be needed during the establishment of the initiative, and grant support continued to fund research activities.

Data-sharing initiatives created to inform disaster response or those that seek to promote wider access to datasets are less likely to focus on revenue models, whereas **initiatives that have a longer operational time period or the ability to 'close' their dataset, have the ability to explore product or service models.**

Articulating the value of data sharing is crucial to attracting funding or building revenue models. All data-sharing partnerships need to clearly articulate their purpose and value-add to achieve long-term financial sustainability.

6 High-quality, interoperable data builds the technical foundation of the initiative.

Identifying relevant, reliable data sources and ensuring the quality of their data will determine the robustness of the technical model behind the data-sharing initiative. Initiatives must balance diversity of data sources with the resources needed to clean the datasets and ensure interoperability. Exercises such as data audits or data-gap assessments can help clarify the data needs of the initiative.

The resources invested in data cleaning depends on the expected data use and the platform that the initiative is trying to build. Initiatives that are looking to build data exchanges or open datasets tend to spend more resources on ensuring interoperability, whereas those that seek to inform

implementers or governments can spend months cleaning datasets and ensuring adequate data quality. Clarifying metadata, standardising data formats, and implementing standardisation processes at the beginning can be resource-intensive, but it forms the foundation of the technical output of the initiative.

7 Big data analytical models need to be supported by ground-truthing survey data, qualitative interviews, contextual experience, etc. Without this, such models run the risk of causing harm.

New data analytical methods present an opportunity to utilise data sources quickly and to more cost-effectively identify patterns and inform policy and aid decisions. However, concerns exist on the representativeness of these new datasets, and consequent bias in the analytical models and decisions informed by them. **Bias can be exaggerated both by a lack of representativeness in the dataset and a lack of diversity in the teams building the models.**

Data-sharing initiatives reviewed for this study adopted a variety of approaches to identify and overcome limitations in their analytical models. Detailed data limitation sections; use of multiple data sources; development of technical sample bias correction methods; verification of representativeness through comparison with socio-economic surveys; household survey data to support analytical models, qualitative interviews and others were used in different use cases. More research on the effectiveness of different verifications exercises will be useful in developing robust models.

Interviews suggest that measures such as inclusion of national and local team members in the data science team are important for the identification of bias in datasets and reducing the bias of analytical models.

8 The breach of data security and privacy present legal, reputational, and data risks. Depending on the nature of the data shared and the data-sharing agreement between partners, access to raw data and data-based insights could be tightly regulated.

The ability of an initiative to implement the access terms agreed upon by partners is essential to its credibility and sustainability. Any breach of agreed-upon access norms can bring huge legal and reputational costs to the initiative. Thus, it is crucial that partners trust the operational protocols of the platform to adequately guard against such breaches.

Adherence to robust data-security protocols is a crucial element of the technical infrastructure of a data-sharing initiative. Initiatives can face security risks from data breaches and re-identification of aggregated, anonymised data. **Exercises such as regular security reviews and audits, privacy-impact assessments, adoption of internationally accepted protocols to minimise de-identification risks, etc. must become a part of operating procedures within data-sharing initiatives.**