Discussion Group on Data and Data Driven-Technologies

Summary note

Chair
- Dr. Cathy Mulligan, Panel Member

Moderator
- Amandeep Singh Gill, Executive Director & Co-Lead of the Secretariat

Guiding Questions

The meeting sought to address the following questions:

- What does cooperation between stakeholder groups, and across borders, currently look like around the management of personal and other types of data? What are the key elements of successful cooperation? Who or what are the enablers of cooperation? What seems to be lacking?
- How can privacy, security and other public interests be ensured, while also supporting the development of new data-driven technologies?

Executive Summary

Participants highlighted examples of inter-sectoral cooperation and collaboration around data, and identified the main challenges and barriers. The discussion focused first and foremost on the issue of aligning incentives. Second, participants discussed the challenges that data privacy and data fragmentation pose to enhanced data collaboration between sectors and sometimes even within governments (between agencies or ministries or between national and state levels). Several participants pointed to the new challenges brought on by a data-saturated world, the need for a better understanding of the political economy of data, and emerging data ethics questions. Rather than thinking of data as “the new oil” and conceptualizing ‘data ownership’ as a zero-sum game, it could be beneficial to think of data as an ecological system, which can include certain public-private ecologies. Lastly, three “layers” were identified that enable (or hinder) data collaboration: the “hardware” of regulation and laws, the “middle ware” of standards and architecture, and the “software” of ethics and trust building between individuals.

The meeting followed a pilot virtual discussion on data held in September 2018 and preceded a consultation on data held in New York in partnership with the UN Foundation on 30 November 2018.
Meeting Summary

**Big Data for social good**

Most examples of leveraging big data insights for “social good” or non-commercial purposes have involved partnerships and collaboration between private sector companies and government agencies, international organizations, and/or academic institutions and non-profit organizations.

Big data analytics derived from commercial activity, such as e-commerce, can reveal useful insights for social development measurement, monitoring and policymaking. For example, trends in Alibaba’s e-commerce activity data show specific regional differences and gender differences, which are useful indicators of social and economic development. Additionally, mobile phone data from mobile network operators (MNOs) has been analyzed to reveal insights about transportation and map the spread of diseases. Financial transaction data has been used by non-profit organizations seeking to enable financial inclusion.

*Lessons learned from “mobile data for development” initiatives*

Several participants (including representatives from private industry, academia and research organizations) described specific examples of data collaboration projects that generally involve one or more MNO and either a government or international organization.¹

The biggest challenge around data collaboration with outside organizations is data privacy. Organizations like MasterCard and Telefonica have to anonymize data before providing insights to third parties and other stakeholders, and this requires technical expertise and understanding of data. Yet some companies – and some governments – lack understanding of security and anonymization solutions. There is need to increase security awareness across all organizations involved.

Some companies and organizations are trying the “Open Algorithms” approach, whereby data stay in the organization where they originate, and standardized algorithms are created for certain projects (defined by committee). Within the context of a partnership or collaboration, once the third party makes a request, the company runs the algorithm on the commercial data in-house and shares the insights.²

*Fragmentation and “pilot-itis”*

Success happens when projects are implemented on an ongoing basis (e.g. mobile data regularly being used for health or traffic metrics), not just for research or piloting/testing. To get to this stage, incentives or sustainable business/operating models are needed. These do not currently exist.

An additional hurdle is that if data needs to be combined across industries, standards need to be harmonized. Analyzing data from different companies in different countries could otherwise result in wildly different outcomes. This presents questions of

---


² “Open Algorithms” project: https://www.opalproject.org/
“scalability,” particularly in the quest to use big data for SDG measurement or monitoring.

Participants pointed to an effort under way within the GSMA, the mobile industry association. It is a first-of-its-kind effort to spur industry-wide cooperation to agree a set of principles and standard processes for data sharing between companies and public sector/international organizations.³

Incentives

Humanitarian organizations, governments, and companies have different incentives which are not aligned. Participants agreed that alignment of incentives is critical for enabling greater cooperation in the context of leveraging data. As one participant from academia noted, “effective collaboration will require the alignment or the correct chaining of different interests (tech community, researchers, policy makers, individuals etc.) as well as time scales to deliver ‘results’.” Participants presented different perspectives on how to enable alignment.

Taking the example of mobile operators, representative from LirneAsia noted that incentives are different for different stakeholders: mobile networks are often used in developing countries as tax revenue collectors. They therefore have an incentive to show the amount of public good they are doing in relation to using “data for development.” Additionally, new R&D and derivation of insights on their own data is valuable for improving their own business. For governments, incentives for cooperation with MNOs revolve around regulatory issues. For the users of big data insights (which may include governments, international organizations, research or civil society organizations), incentives are varied. A private sector representative noted that the incentive must be more concrete than just the “public good”, as this is not compelling enough to make all actors work in the same direction. It is already difficult to align all the stakeholders within an organization to advance data science work: it is even harder to do so between organizations.

A participant from academia noted that the timelines and incentives of the research environment are very different from those of government. Research tends to work on longer timelines, while government officials may change jobs/departments or need to move onto different project portfolios. One starting point for alignment might be to at least make sure that key questions are formulated appropriately for different stakeholder groups. While the motivation or incentives of different stakeholders may be different, they can still work together if all are seeking answers to the same set of problems.

³ A recent report attempts to outline the business case for scale: https://www.gsma.com/betterfuture/resources/sustainable-business-models-report
Ethics and ownership (“whose data is it anyway?”)

The entire process around data from collection point to the end (interaction human/machine and machine/machine) raises ethical concerns ranging from who should have access, who should hold what amount of data, and in the context of citizen rights - ownership and control - who derives benefit from data. As a starting point to tackle these questions, a participant from academia proposed a few ways to develop digital cooperation using 'ethical-by-design' approaches: 1) ensuring ethical behaviour throughout the design processes, 2) actually 'engineering for ethics' to ensure ethical behaviour of the systems themselves. Participants agreed on the need to develop codes of conduct and culture development for both aspects.

Recognizing the political economy of data

A representative from an India-based NGO, noted that we need to work on institutional development and a re-framed understanding of data ownership. Today, we often speak about personal data protection (GDPR describes the “data subject”: who the data is about). However, data also belongs to a collective - groups, communities and nations. The question of ownership is thus not just individual. Does commuting data in a city like Geneva belong to the commuters? If this data can be useful for improving traffic flows, what kind of rights do those commuters have about how their data is used? We need a cooperative framework to lay out who can do what with data resources. A mix of public-private ecologies around data needs to evolve. For example, in the UK there is already discussion about ‘data trusts’, the French AI strategy talks about ‘data infrastructure’, and India is developing data infrastructure and ‘data lockers’ for collaboration, say around health and financial inclusion.

Educate individuals, shift the power dynamics

In a similar vein, representatives from IEEE reminded that education of individuals is needed to unlock new modalities of data collaboration such as data exchanges. If individuals have access to data (e.g. patient data), this shifts the power dynamics. Today, stakeholders aren’t incentivized to do this. Today doctors give notice to patients on what to do, based on the data they have about their patients. But why not allow patients to actively manage their own health data?

Data flow & data localization

On the broader policy layer, national laws and regulation related to international data flow and data localization were noted as another dimension that is critical to cooperation.

Representatives from the financial services industry noted that the patchwork of privacy laws that exists today creates a difficult and complex legal framework. This legal complexity (with sometimes divergent, or even conflicting laws) makes it difficult to cooperate. This was echoed by a representative from the mobile industry who mentioned they are therefore in support of global free trade (including digital trade and ecommerce) such as via ICC and WTO dialogues, and its objective to build a network or patchwork of modern, high-quality trade agreements focused on services. “Data

---

4 This summary paper provides an overview of the emerging ‘digital ethics’ field from a system design and engineering perspective: https://zenodo.org/record/1481834#.XB2ugfxOlp-
flows are the backbone of the digital economy and crucial to boosting growth in all sectors.” Companies need the continued ability to transfer data across borders, while respecting applicable data protection laws and enhancing trust. Data localization obligations hinder collaboration. Data protection laws should form a common architecture of core principles across key nations (the EU GDPR has proven to be a foundation) and provide workable transfer mechanisms, which would enable companies to transfer data cross border in full compliance with the law.

Some solutions: technical, standards, principles, and the human dimension

The human dimension: it’s about people
Building relationships with individuals was cited as critical when seeking to build data collaboration with government. Statisticians and data “technicians” in national government entities are a key entry point. At the same time, it is necessary to simultaneously cultivate relationships with top-level government officials (e.g. Ministers, heads of agencies) to secure commitments, socialize new ideas or processes, and find ways to let them encourage and even champion their technical colleagues to try new things, and to avoid resistance later.

Technical solutions
A participant from a software company that provides a platform for data engineering, data warehousing, machine learning and analytics, noted that they are experimenting with technical solutions to address the challenge of sharing data across organizations, and explained an effort at “federated learning”, whereby data doesn’t have to be in a centralized place.5

Principles
Data for Democracy provides a platform where data practitioners from different industries and countries work together on sets of problems for social good. This does not solve the policy and regulatory issues but it begins to set the cultural framework for consensus building among stakeholders. One of the first questions they ask stakeholders is “what are the norms that you are using to determine what-is-ethics or what-are-the-values”.6

Standards
IEEE promotes silo breaking and cooperation by using a paradigm of standards development that is inclusive and is consensus based. Everyone who has a vested interest and something to contribute is welcome to the table. This generates cooperation not only across borders, but importantly across disciplines (not just sectoral: industry, policy, regulator).

Safeguards
In some scenarios it is preferable to be highly conservative about data collaboration and focus first and foremost on safeguards. One large humanitarian organization, for instance, is focused on risks as they work to protect people in conflict and in situations of violence. Based on research, they have identified a range of misuses of metadata (e.g. messaging apps, cash transfers, digital identity systems). There are risks in

5 https://vision.cloudera.com/an-introduction-to-federated-learning/
6 https://www.datafordemocracy.org/project/global-data-ethics-project
collaboration, for example, biometric or identity data that is shared across organizations could end up in databases of law enforcement agencies in repressive governments or in countries engaged in active conflict. Thus, they share data with other agencies only if there is a clear benefit to their beneficiaries, and if the use case meets a humanitarian purpose. A few criteria for cooperation that need to be borne in mind are: Who do we empower ultimately? Do we have a clear purpose before we collect and share data? Is that common purpose humanitarian or other?

From a procedural point of view, they need a clear agreement on the nature of the cooperation at the beginning, and conduct a Privacy Impact Assessment every time they want to share or collect data with an external or third party.
Data matters. Breakthroughs in machine learning, and the availability of computational power, cloud computing and data at low cost have accelerated the pace at which patterns found from data, particularly large multi-variable data sets, can provide insights which may inform, assist, or replace human decisions in myriad of domains. This has fueled new industries, innovations and applications in an increasingly diverse array of fields, including robotics, self-driving cars, health, defence, finance, consumer services, advertising and retail, government services and other areas. Indeed, both the volume of data and the development of data insights appears to be accelerating.

This has also given rise to profound opportunities as well as a plethora of issues including:

- concerns that big data and related technology may worsen the digital divide;
- calls for increased data sharing and treatment of data as a public good in certain contexts;
- regulations on the flow and storage of personal data across borders and jurisdictions;
- challenges for governments around the world to understand and derive benefits from data;
- deterioration of the public’s trust of private sector guardianship of user privacy and personal data security;
- the risk of perpetuating unfair social biases in data-driven systems;

Given the rapid pace of technological developments, traditional forms of regulation might be insufficient to handle the social, economic, policy, regulatory, and governance needs brought on by data-driven technologies. Alternative multi-stakeholder approaches and cooperation mechanisms may be necessary. There is an opportunity to learn from use cases (e-commerce, finance, health, humanitarian, scientific research, development to name a few), to improve incentives for data sharing and cooperation, build capacity to enhance participation of all stakeholders, and promote safety, security and the legitimate rights of users.

In reflecting on the numerous issues pertaining to data and data-driven technologies, this discussion group will cover the following questions:

1. What does cooperation between stakeholder groups, and across borders, currently look like around the management of personal and other types of data? What are the key elements of successful cooperation? Who or what are the enablers of cooperation? What seems to be lacking? What is best practice in breaking silos between stakeholders?
2. How can privacy, security and other public interests be ensured, while also supporting the development of new data-driven technologies? Which type of cooperation is required to create an enabling environment for preserving this balance?
3. How can policymaking processes, and the capacity of governments or international organizations to address data related issues be improved? What role do industry, government and civil society need to play?
4. What principles or values should underpin cooperative approaches to designing and developing technology, policy, and regulatory solutions that are inclusive, responsive to technological progress, and/or tailored to the context in which they are being implemented? How can stakeholders better manage the benefits and risks of data-driven technologies?

5. How can the potential of data and data-driven technology be used to support progress towards achievement of the **UN Sustainable Development Goals**? What is best practice in the area of data and development – and what methods of cooperation have been found effective in this arena?