Data for
development:
the road ahead

FOREWORD

More than ever, data is a crucial element shaping our societies.

Traditionally, "data" refers to facts and statistics collected together for reference or analysis. In the context of international development, broadly, data helps policy makers and citizens track progress towards a development goal or objective, aid in decision-making about public services, improve resource allocation, and help to map gaps.

Today, the exponential growth of information technologies and systems have created massive troves of data and it is increasingly shaping every aspect of our lives. It is estimated that 90 percent of the data in the world was generated over the last two years alone. This "data revolution" has the potential to help make marginalized communities and their needs become more visible to policy makers, inspiring new solutions to local and global challenges.

Yet, we know that most developing countries lack the infrastructure and skills to collect, house, analyze and share data to enable their citizens and civil society to advocate for equitable development pathways. In a world shacked by the COVID pandemic, we all see that clearly shows the fundamental need to have strong data systems as a foundation for fighting the devastating spread of the disease as well as it is extremely unequal social economic impacts.

As a funder of research for development, data is at the core of IDRC's mandate. We see data and evidence at the centre of innovative processes to scale sustainable solutions for development. Innovations around data can help to create enabling environments where new solutions can flourish.
For instance, mature lines of IDRC's investments on data for development focus on encouraging governments to establish open data policies so that vulnerable populations can begin to understand and use data to generate evidence

to support advocating for themselves, and establish civil registration and vital statistics infrastructure to increase access to basic rights like health care and land ownership for vulnerable or transient populations.

We also have seen that data can be a key element of more inclusive governance and better functioning democracies. In a society increasingly driven by data, we need to continue to investigate how data is addressing inequalities or causing other inequalities. At the same time, we also know that data systems need to be considered carefully. Data has been used by authoritarian regimes using telecom and internet data to profile and persecute groups that lack the skills and technology to protect themselves. Well-funded special interest groups have leveraged data and data systems to manipulate behaviour and polarise public debate, leading to downstream effects that exacerbate inequalities. We need to ensure the creation of accountable, responsive, and transparent institutions that can protect the privacy of citizens while mobilizing data to improve the lives of the poor.

This report "Data for Development: the road ahead" describes the discussions held at an workshop in Montevideo in March 2020, an initiative that we co-hosted with ILDA and counted with the participation of many partners that came together to build bridges between different silos and explore a more systemic view of the data society from a Southern perspective. It is also part of our renewed commitment to support a research community in the global South that will drive new data-driven innovations, better governance and a more efficient infrastructure in the next decade. We need to support the creation, collection and availability of key development data, making sure that we help to develop this core infrastructure in a way that it recognizes vulnerable populations and their needs.

This report is a contribution to the development of this broad agenda exploring how to leverage data to better respond to the immediate crisis triggered by the COVID 19 virus as well as the main challenges that will continue to shape our societies in the next decade. It is also an invitation to join this emerging community that aims to continue to build the capacity of actors to identify, collect, use and transform data into better decisions and better lives. We look forward to continuing to build this agenda together!

Fernando Perini
Regional Director, Latin America and the Caribbean

Credits

This report is the outcome of the collective input of the following colleagues who attended or participated our workshop in Montevideo, Uruguay:

IDRC coordination: Fernando Perini and Katie Clancy

Report coordination: Fabrizio Scrollini

Report authorship: Fabrizio Scrollini, Katie Clancy, Silvana Fumega and Ana Sofia Ruiz

Research support: Maria Esther Cervantes

Workshop facilitation: Fabrizio Scrollini, Silvana Fumega, Katie Clancy, Fernando Perini and Juan Belbis

Visual Facilitation: Sofia Donner

Edition: David Schwartz

Graphic design: Daniel Araya Vega - www.indefinido.com

Workshop participants: Alexandrine Pirlot de Corbion, Alison Gillwald, Arturo Muente, Christian Perrone, Christopher Tuckwood, Eva Blum-Dumontet, Federico Lecumberry, Florencia Serale, Florencio Ceballos, Francisco Mata Chavarria, Fredy Rodríguez, Gonzalo Iglesias, Grace Bomu, Helani Galpaya, Jamila Venturini, Judith Mariscal, Khairil Yusof, Laurent Elder, Leonida Mutuku, Lucía Abelenda, , Mariana Fossatti, Maurice McNaughton, Natalia Carfi, Priscila Chaves, Pyrou Chung, Ramiro Albrieu, Renata Ávila, Steve Walker, Tim Davies.

1. Introduction:
Data for Development

The last 20 years have seen a rapid evolution of the "data for development" agenda. The Data Revolution report (2015) defines this agenda in several different ways:

> The data revolution is an explosion in the volume of data, the speed with which data are produced, the number of producers of data, the dissemination of data, the range of things for which data exists, and the demand for data from all parts of society. These data come from new technologies such as mobile phones and the "internet of things," and from other sources, such as qualitative data, citizen-generated data and perception data.

The data revolution for sustainable development is:

- the integration of these new data with traditional data to produce high-quality information that is more detailed, timely and relevant for many purposes and users, especially to foster and monitor sustainable development;
- the increase in the usefulness of data through a much greater degree of openness and transparency, avoiding invasion of privacy and abuse of human rights from misuse of data on individuals and groups, and minimizing inequality in production, access to and use of data;
- ultimately, greater empowerment of individuals, better policies, better decisions and greater participation and accountability, leading to better outcomes for people and the planet.

After two decades of pilot studies and research, data have been used in several innovative ways to enhance an inclusive development agenda, including to achieve

the 2030 Sustainable Development Goals[1]. Data is being used to predict floods, improve agricultural value chains, transform public transport, deliver better and more inclusive public services and improve government transparency.

Governments and other institutions have been collecting and generating data about their citizens, economies and societies through censuses and other instruments for thousands of years. However, the advent of the internet, proliferation of the mobile phone, and new sensors and sources of data as well as the digitization of governments mean that more data has been produced in the past two years than ever before. Advances in computing and analytics, in particular artificial intelligence,[2] have the potential to amplify the value of data through better prediction, pattern recognition, and automation. The range of data producers has increased and now includes not only traditional producers such as national statistics organizations, government departments, research organizations, civil society organizations and private sector organizations but also user-generated data compiled by global communications and social media companies. However, there has been little assessment of whether these advances improve efficiency or of the potential harm they pose in different contexts. The need for a shared framework to identify benefits, harms and challenges according to context seems well established (Smith and Neupane, 2018).

The questions of what data are collected, how, by whom and for what purposes takes on a new dimension when data can be combined and used in different ways thanks to digital technologies and information systems. Countries in the Global South often lack access to data collected about them, or must privilege international agendas over their own priorities. While more data are available than ever before, particularly in the Global North, there remains a lack of disaggregated data about key development issues. This dearth of disaggregated data for development, in combination with large data gaps in the Global South, disadvantages vulnerable and marginalized communities, such as people living in extreme poverty, ethinc, indigenous and religious minorities, informal migrants and refugees, women and minorities. There are several potential explanations for this situation. Global South countries often

1. In particular, we highlight the following reports: A world that counts (United Nations, 2015), the State of Open Data (Davies et al., 2019), the Open Data Barometer (World Wide Web Foundation, 2018) as well as the upcoming World Development Report (World Bank Group, 2020)
2. For the purposes of this paper, we use Vinuesa et al.'s (2020) definition of AI as any software technology with at least one of the following capabilities: perception—including audio, visual, textual, and tactile (e.g., face recognition), decision-making (e.g., medical diagnosis systems), prediction (e.g., weather forecasting), automatic knowledge extraction and pattern recognition from data (e.g., discovery of fake news circles in social media), interactive communication (e.g., social robots or chat bots), and logical reasoning (e.g., theory development from premises) https://www.nature.com/articles/s41467-019-14108-y

We thus face a fundamental challenge: looking at data solely as a resource ignores the ecosystem that needs to exist for data to truly be catalytic for development.

lack the infrastructure and capacity to generate data and put it to use. Furthermore, user-generated data are often in the hands of global corporations outside these countries' jurisdiction. Quality, completeness, access, and degree of ownership are all factors affecting the use of data in Global South countries.

Is data "the new oil?": Moving beyond extraction

The promise of data-driven economies and societies has led government officials, experts and activists to refer to data as "the new oil." This analogy implies that data is now one of the most valuable resources that countries can use to achieve sustainable development. While it is true that data's economic value has increased, this increase is a direct result of the greater number and affordability of analytical techniques that use data to make more accurate predictions (which, in turn, can improve decision making, automation and more). The data-is-the-new-oil analogy envisions data as the key resource in a value chain where information is extracted, refined and turned into something more valuable (though some take issue with this framing from an economic standpoint). The analogy also potentially offers us a cautionary tale about the pitfalls of extractive industries in terms of wealth concentration, equality, sustainability and impact on the environment and livelihoods.

Companies based primarily in the Global North—but with global reach—developed a successful business model that led to a major concentration of personal data, which increased the companies' wealth and influence. Through specific projects, these companies have shown how the data they collected in combination with public data and other types of data have led to innovative uses that can advance understanding of development challenges. Furthermore, the use of artificial intelligence (AI) techniques is evolving at a rapid pace, transforming different sectors such as agriculture, manufacturing and government services, which, as Smith and Neupane (2018) argue, has prompted significant questions about fairness and inclusion.

We thus face a fundamental challenge: looking at data solely as a resource ignores the ecosystem that needs to exist for data to truly be catalytic for development. If data are to meaningfully contribute to development, how data are managed, governed and used becomes fundamental not just for economic gain, but also for ensuring rights and good government and for maintaining progress towards inclusion, equity and equality.

3. Agrawal, Gans and Goldfarb (2019). Prediction Machines: The Simple Economics of Artificial Intelligence,
4. "Why data is not the new oil." https://truthonthemarket.com/2019/10/08/why-data-is-not-the-new-oil/amp/

By "data governance" we mean a set of legal, social and technical arrangements about access, control, transparency and rights over data.

This report is being produced in the midst of a global pandemic during which governments, private sector companies, civil society organizations and international organizations are trying to make sense of the data they need to face the pandemic. However, they face several challenges; health system data is not up to date in several countries, a problem further hampered by the multitude of platforms and lack of interoperability that prevents health systems from adequately planning for this contingency. Scientific data (e.g. virology samples) are now shared more widely but open collaboration and data sharing remain complex tasks. Companies and federal and local governments are rapidly trying to develop their own versions of contact-tracing apps with several unanswered questions about their design, privacy implications and effectiveness. Open and reliable data about the pandemic remains difficult to find given the current way global communications platforms operate and there have been some dubious uses of the COVID crisis to justify expanding or creating new covert—or even illegal—surveillance systems.[5]

Some previous efforts, such as the International Aid Transparency Initiative (IATI), launched in 2008, have been central to "establishing the methodologies and platforms needed to provide open aid data within the Global South and emerging market economies" (Weavell et al., 2019) and other initiatives such as the UN OCHA's Humanitarian Data Exchange (HDX), provide step-by-step guidance for sharing data while adhering to strict practices of organizational and individual accountability. According to the 2019 State of Open Data report (Weavell et al., 2019), as of March 2018, there were over 6,500 datasets and hundreds of participating organizations sharing a wide range of open data, including assessments, geospatial data, population data, and more. However despite these efforts, the world remains far from achieving the effective use of data in emergency contexts.

The current state of technology should allow for a better response and the development of tools that enhance data openness, data sharing and the safeguarding of fundamental human rights. The cause of the deficiency seems to be more a problem of data governance, and of the rules and incentives for collaboration to facilitate it, than of the technology needed to confront these challenges. By "data governance" we mean a set of legal, social and technical arrangements about access, control, transparency and rights over data. It is also evident that we need to address the complex and long-standing challenges in terms of inclusion, governance and fundamental rights in order to fully embrace the digital opportunities in times of need such as the current pandemic.

5. See, for instance, the work Privacy International on the matter available at ttps://privacyinternational.org/campaigns/fighting-global-covid-19-power-grab

About this report

This report is the result of the Data, Artificial Intelligence and Society workshop that took place in Montevideo, Uruguay in March 2020 and was co-hosted by ILDA and IDRC. The workshop brought together a diverse group of experts to discuss the current state of the data-for-development field. A number of participants who were unable to travel joined by videoconference. This report presents a synthesis of the main points articulated in the group discussions of the workshop sessions. Where possible, the report provides a set of questions, issues, and cases which the group agreed would be important to raise or highlight in the future. The workshop also employed a scenario-building methodology to engage the participants in a discussion of how the field may evolve. The report offers reflections on the way forward in terms of research questions and actions that like-minded research or activist organizations could tackle to engage with the emerging new landscape of data for development. Finally, the group used a visual design methodology to capture the main points of the debate[6] as well as to facilitate the scenario building exercise.

6. The report does not attribute statements to specific participants except when they have consented and provided a specific quote. Discussion groups were run under Chatham House rules for attributing comments. Individual presentations are attributed in our graphic work.

2. The big picture

Digital technologies are now pervasive in all activities across the globe and, as a result, a large amount of data is produced daily. The data collected concerns all aspects of human activity that have been digitized. For instance, population movement data can be obtained from cell phones which are increasingly and cheaply available in the Global South as well as the Global North. Weather data from satellites and sensors allows us to understand more about climate change and help farmers to adjust and prepare to improve crop production. Everyday, millions of people provide feedback about the quality of public services through several online platforms that, ideally, can then be used to improve those services.

In theory, the billions of zettabytes of data in the world could be used in service of solving several public problems. In practice, the data is fragmented, often poorly analysed and often ignored by decision makers. The problem is now evident amidst the chaos of (dis)information the world faces during the COVID-19 pandemic, where decision makers are struggling to find key facts about how health care systems operate, allocations of public contracts and resources, clinical data, and data sharing with international partners. Furthermore, data are often collected with little regard to human rights, including privacy of individuals, as several scandals and cases have shown in recent years.

However, in the last 20 years there has been an evolution as well. The acknowledgement of the importance of increasing access to and sharing data has gained strength in several forums. Privacy and human rights issues can no longer be ignored by eager technology enthusiasts in governments and the private sector as something that will eventually be dealt with rather than as a fundamental pillar in the development of technology interventions. Also, as more data becomes available, new tools based mostly on machine-learning techniques (AI) become important tools for automating

analyzing and—potentially—predicting developments in several fields. Taking this evolution into account, the workshop group discussed several issues and concerns about the current environment. The general consensus is that while there are several reasons for the poor use of data to address development issues, a key aspect that impacts the poor use of data in development is data governance. As new tools become available and data becomes more granular, there is increasing ability to track individuals, new opportunities to rank, score and classify them and, thus, to automate decisions based on these classifications. It is therefore important to discuss the potential for autonomous systems to emerge and assume critical societal functions. Governing the emergence of these technologies is a difficult task.

Furthermore, as in other policy fields, policy transfer from the Global North to the Global South may be inefficient at best or, at worst, may be a new manifestation of "colonialism." In short, governance discussions are often about choices that societies need to make about power, inclusion and the different ways in which these emerging data and technology scenarios will affect them. While regulation harmonization is always desirable in an increasingly networked world, there has to be room to critically examine and adopt adequate regulations that take into account regulatory context. There is a recognition that there is a need to harmonize regulation, innovation and human rights considerations as in other fields.

Governance issues require engagement with different stakeholders to understand the incentives they have to contribute towards the use of data and AI for development. Private sector entities such as platforms hold a trove of (mostly personal) data that could be used to inform public policies when combined with official data such as that produced by the statistics agencies. The Sustainable Development Goals (SDGs) and the Data World Forum already provide examples of early and potentially useful collaboration. Another example of data collaboration is a project carried out by Facebook researchers with Columbia University's Center for International Earth Science Information Network. Bonafilia, Gill, Kirsanov and Sundram (2019) explain that their goal was to provide reliable and precise maps[7] that are generally not available and the project produced, among other things, density maps for the entire population of Africa. While these projects are useful, it is important for data collaborations to discuss clearly what is being shared and how it is being shared. The idea of "data trusts" or sharing agreements has been around for a while, but it seems there is not enough technical and political will

---

7. The maps referenced in this project can be found on: https://data.humdata.org/dataset/highresolutionpopulationdensitymaps

> Transparency and accountability are essential for the citizenry to understand how these systems work, but even with these measures in place, decisions may not be fully understood (Annani and Crawford, 2018)

to make this sort of collaboration happen at the moment, at least when using data for the public good.

There are risks to sharing data under these new circumstances. Several civil society organizations, such as the Open Data Institute and the International Data Responsibility Group (Taylor, 2016), have advocated for treating data as a public good, addressing digital inequalities and enjoying a critical engagement with development leaving no one behind. Furthermore, data protection and privacy rights are fundamental in discussions pertaining to data use. We are focusing on leaving no one behind, but the benefits have not been enjoyed equally. Finally, there is the risk of not counting people or vulnerable groups due to technical or political and policy bias, but giving more visibility to these people or groups might inadvertently affect them negatively.

Besides the challenges related to data use discussed above, another important challenge in all scenarios are the unresolved issues around internet connectivity. Universal, affordable access is not yet a reality. Thirty years after the emergence of the world wide web, there remain people who are not able to fully take advantage of the benefits the internet brings in terms of access to knowledge, education, commerce as well as social and inclusion opportunities.

Digitalization and datafication are not silver bullets. Some interventions may make sense while others could adversely affect the desired development outcomes. In addition, using AI technologies to make government decisions increases the complexity associated with automated decision making by amplifying existing biases. Also, AI technologies require highly specialized knowledge and infrastructure and might be poorly executed without a full understanding of the existing infrastructure and data available in the Global South. Transparency and accountability are essential for the citizenry to understand how these systems work, but even with these measures in place, decisions may not be fully understood (Annani and Crawford, 2018). In worst-case scenarios, the use of these technologies could foster repressive measures against individuals, collectives or entire populations. In the group discussions, there was a consensus regarding the importance of looking beyond data processing and focusing on sectors and specific implications in terms of capabilities, development and human rights to understand what kind of interventions governments and other stakeholders could use.

Finally, a cautionary note about the role of governments in this new scenario. During the first 20 years of the Web and its related activities, there was a general perception of 'benevolent' disruption that — on balance — resulted in more good than harm done to countries in both the Global

South and the Global North. Governments were barriers, lacked the understanding to regulate this new phenomenon and were ill equipped to take significant measures given the global dimension of the regulatory challenge. The use of data and AI for several different purposes — some of them questionable — is no exception. But as the current COVID-19 crisis is showing, governments are still important actors able to exercise power and control: by closing borders, restricting people's movements and using data, AI and telecommunications infrastructure for various ends. Furthermore, limited public understanding of data-related issues as well as the need for data governance to coordinate international responses to pandemics (or other global health issues) still requires governments to be able to engage with stakeholders across multiple sectors. In short, governments will become more central players in the years to come as data are used in a more granular way, directly and indirectly affecting the lives, rights and activities of millions of people.

Key questions

- What kind of technical and governance arrangements are needed to share data and use AI and for what purposes?
- How do we ensure active participation of all relevant stakeholders in the context of the Global South to enable the eventual development and maintenance of these governance frameworks?
- How do we build trusting relationships between the private sector, civil society and public sectors to share data and use data for development outcomes?
- How to embed key principles about human rights, dignity, protection and inclusion of vulnerable groups in this context?
- How do we effectively use evidence in the contemporary policy making context to advance a common data governance and AI agenda ?
- How do we enable a set of diverse, plural, international and relevant voices to participate in international forums to foster cooperation in terms of norms and practices?

Cases

- Data República (CEPEI) working with Telefonica on a data sharing agreement.

Data Republica[8] is a data laboratory for sustainable development in Latin America created through a partnership between CEPEI and Teléfonica. Their mission is to strengthen data ecosystems and the mapping of information sources to generate new knowledge. The platform collects and centralizes data from different institutions and associates them with sustainable development goals. Phase 1 of this project operated in Mexico, Costa Rica and Colombia collecting approximately 1300 datasets related to the sustainable development goals with the collaboration of public and private actors. This project also delivered data journalism training in collaboration with organizations such as Social Tic, CONNECTAS, Miríada and Escuela de Datos. The final part of this project facilitated the publication of stories and visualizations based on the datasets published as part of the project.

In a tweet

"When using non-traditional data sources it's important to keep in mind that tech is not always the answer, the answer is how you create relationships and trust with the owners of the data." **- Fredy Rodriguez**

"It is important to highlight our lack of data to understand the impacts of datafication in society." **- Alison Gillwald**

"We need to rethink the way that we articulate evidence and use data in our current context, and make an effort to understand the legacy of inequality that data reflects."
- Fabrizio Scrollini

8.   See https://www.datarepublica.org/acerca.

3. Civic uses of AI and Data

One of the early assumptions about open data is that it would create social, economic and civic value (Eaves 2012, Davies et al., 2012). This view was anchored in the powerful idea that if more data and more information were available, it would be used by societies to create more knowledge. This knowledge would theoretically lead to the reduction of several asymmetries of information and lead, in turn, to greater accountability, efficiency and effectiveness. In the current state of the world, these assumptions may seem naive. Nevertheless these assumptions fueled, and, up to a point, delivered evidence that data use and transparency do bring about changes in sectors such as health, welfare, finance, extractive industry, land ownership and anticorruption (van Schalkwyk et al., 2017, Davies et al., 2019).

Artificial Intelligence or AI comes from a different perspective. The field derives from the work of a British mathematician, Alan Turning, who in 1950 posed a simple question: Can Machines think? Seventy years on, the answer remains elusive and the field is still evolving. The field underwent several phases, including the so-called "AI winter" from the 1980's to 2000 when the field went almost dormant because the grand promises made by its promoters went largely unfulfilled (Nilsson, 2010). For the purposes of this paper, we use Vinuesa et al.'s (2020) definition of AI as any software technology with at least one of the following capabilities: perception—including audio, visual, textual, and tactile (e.g., face recognition), decision-making (e.g., medical diagnosis systems), prediction (e.g., weather forecasting), automatic knowledge extraction and pattern recognition from data (e.g., discovery of fake news circles in social media), interactive communication (e.g., social robots or chat bots), and logical reasoning (e.g., theory development from premises).

AI's application to civic uses is not as evident as is the open data field. The AI field has found application mostly for

> Nevertheless, there is reason to believe that openness in terms of data, algorithms and processes will yield civic, social and economic value.

narrow public problems or for issues such as health systems or traffic. Nevertheless, as Davies (2019) reports, there are several instances of AI being used for the public benefit. Examples include Serenata de Amor, a bot that monitors MPs expenditures in Brazil, or the use of decision trees to influence public policy around dengue in Paraguay (Pane et al., 2015). Several specialized centers across the world are working to increase public sector uses of AI. For instance, the MILA Center in Canada is currently organizing an "AI on a social mission" conference[9]. Such activities are recent, however, and remain exploratory. Nevertheless, there is reason to believe that openness in terms of data, algorithms and processes will yield civic, social and economic value. This is reflected in the interest shown in this work by funding organizations such as the Rockefeller Foundation (Sha, 2020).

The group concentrated on discussing examples of various practical uses of data and emerging technologies in different environments. A common thread among the examples discussed is that often organizations or groups of involved citizens took advantage of available data to make decisions or problems visible. Some examples discussed dealt with the gender pay gap (O'Donnell et al., 2020), femicides (Fumega, 2019) or corruption data (Florez and Tonn, 2019). The increasing availability of data will reveal its inadequacy and the biases that influenced how it was initially gathered and constructed, which will inform our understanding of the problem. Additionally, as data become more widely available and its use is facilitated, policy makers may avail themselves of it and, one hopes, make better decisions. Nevertheless, not all organizations operate in the same data and policy environment. This presents an opportunity to discuss the potential of data coalitions as a way to coordinate around topics that benefit from a mix of open and shared sources.

Data standards matter if they are considered part of the technical infrastructure that enables interoperability and usage. However, standardization alone is not a silver bullet that can solve problems by itself in the absence of data. Standardization allows for a sociotechnical and oftentimes political process to emerge identifying key aspects of data and usage, which could be used, for instance, to fight corruption in cases of budgets, contracts or beneficial ownership. Standards and metadata are also important for establishing credibility and providing sources in a context where disinformation is rampant throughout the use of several platforms. Furthermore, data standards and, more generally, data quality could be critical if AI techniques were implemented to deliver valuable automatization, analysis and prediction services for citizens and governments.

9. See http://iaenmissionsociale.com/?lang=en

New technologies are used by The Sentinel Project (2020) in 4 different ways: information gathering; information management; visualization; and dissemination and prevention.

During this session, participants discussed the ways in which machine learning offers new insights on a variety of topics, such as the possible uses of citizen-generated data for public purposes and integrating machine learning insights into regulatory frameworks to make them more innovative. An example presented at the workshop is the work of The Sentinel Project, a Canadian non-profit organization that works to prevent mass atrocities through the use of new technologies. New technologies are used by The Sentinel Project (2020) in 4 different ways: information gathering; information management; visualization; and dissemination and prevention. Machine learning is mainly applied in the information management and prevention areas, in the former by developing a database that will help to organize and analyze the information reported from all sources and in the latter by attempting to identify and counter websites that incite hatred, using mobile phone networks to document abuses and warn threatened communities and employing GPS technology to guide targeted people to safe areas. However, human involvement is still required, especially as this work is very context-based and depends on knowledge of the language and culture of the place that is being monitored.

Key questions

- What is "next" in the agenda of openness and use of data for civic purposes?
- How do we build capacities in governments and data users to foster relevant uses of data for civic purposes?
- To what degree can AI enable better transparency, accountability and inclusive outcomes in civic space?
- How can development interests be represented in data standardization processes, and how far does their representation contribute to pro-development outcomes from AI systems using standardized data?

Cases

- Malaysia's sexual assault & crimes open data.

As a result of the Open Parliament program (The Sinar Project, 2020), Malaysian citizens managed to use Parliamentary Answers to identify datasets that a variety of stakeholders, from data providers to users, identified as a priority to be made available or improved. Some of the requested datasets were crime statistics, specifically about bullying, corruption, theft and sexual assault. Of the 58 requests, 35 demonstrated concer about the need for sexual crimes data to be published.

These needs showed that there are multiple stakeholders affected or interested in the publication of sexual crimes data, ranging from civil society organizations to journalists, who want to use the data to protect children and expose unresolved issues.

- Myanmar interdisciplinary work on open data.

A second case presented by the Sinar Project is the experience of an open data collaboration set up by citizens of Myanmar. This case study (Canares et al., 2017) showed that in contexts where government data is not available, open data initiatives can be implemented "from below" by the people who are interested in the publication of the data, instead of the usual top-down approach. In Myanmar, this resulted in the Multilingual Parliamentary Monitoring Project and the data was the result of a coalition between OpenHluttaw or Open Parliament, which worked with the Sinar Project to synchronize and import data from a spreadsheet into their open parliamentary API. The Popolo open data standard was used to consolidate the data which was then imported into the Sinar Project's Open Data/API for Popolo data (Popit) as an API to build the apps and website of the project.

In a tweet

"What we have learned is that we need interoperability of people as much as interoperability of data."**- Natalia Carfi**

---

10. More information on The Sentinel Project can be found on their website: https://thesentinelproject.org/.
11. More information on this project can be found on their website: openhluttaw.info.

4. An inclusive and fair 'algorithmic' deal

As discussed in the previous section, it is essential to understand that the process through which data are generated, classified and eventually used matters. One of the main underlying topics in all the conversations we had about the use of data is the need to be responsible and inclusive when dealing with data production and algorithm design. During the discussions, the group came to a consensus that countries in the Global South often have to deal with colonial legacies in administrative systems, which, in turn, invisibilize a large part of the population or key aspects about them. This is particularly common with data dealing with women, indigenous communities and minorities. By making these populations and their problems visible through data, governments and societies can develop programs and address their needs in more accurate and relevant ways. Nevertheless, depending on the context, increased visibility for vulnerable populations can be a curse as it might accelerate trends towards discrimination and exclusion; in the worst-case scenarios, it might jeopardize their lives. Further, some initiatives target entire populations without their consent, leaving them more vulnerable. The group agreed that more research is needed regarding the production of data and the design of algorithms, to make sure that certain communities and marginalized groups are not left behind, and are also not exposed to greater damages. It was agreed that AI could replicate existing biases in our societies. There was an explicit call in this section on the need for evidence about how these systems are being implemented in the Global South and their impacts in terms of discrimination and violation of fundamental rights.

In short, a new inclusive, feminist approach to data and AI could foster a more robust discussion around the development of AI in the public space.

The group agreed on the need for a multidisciplinary and interinstitutional approach to understand how AI works for development and the merits of including a feminist approach to data[12] and AI. The group agreed that ethical approaches and legal approaches, particularly in terms of human rights, differ substantially in terms of nature and enforcement. An "ethical" approach does not mean simply to satisfy international human rights standards. The group discussed how an ethical approach could foster inclusion of all persons, visibility of problems and situations affecting not only women but marginalized groups, as well as a more transparent approach to algorithmic decision making. In short, a new inclusive, feminist approach to data and AI could foster a more robust discussion around the development of AI in the public space.

Lastly, the group agreed on the importance of providing quality education systems at all levels to every human being without distinctions. A clear example of this is the need to foster STEM education for girls in order for them to be included in all future conversations regarding technology and data. This is key to developing critically informed citizens who are able to understand and discern the power of algorithms in society.

12. For more information on a feminist approach to data, we recommend looking at "Data Feminism" by Catherine D'Ignazio and Lauren F. Klei., This book suggests alternative ways of thinking about data science and data ethics informed by the ideas of intersectional feminism.

Key questions

- Can algorithms in the current governance status quo become inclusive?
- What is the role of international organizations in replicating Global North/South divisions? What is the role of international organizations in lessening Global North/ South divisions?
- What does an inclusive feminist data approach look like and how could it be achieved?
- How can education systems impact the ethical implications of AI?
- What other models for development are possible that respect human rights and go beyond solutions based on intensive data collection from vulnerable communities in the Global South?

Cases

- ILDA femicide data standardization project.

At the beginning of 2017, ILDA began an exploratory study — supported by IDRC and Fundación Avina — to understand how changes to the production and use of data might contribute to understanding and ultimately the prevention of femicide in Latin America. An action-research methodology was designed to assess the problem, to understand how working with data — particularly open data — could contribute to a solution, and to establish recommendations for the countries involved. ILDA developed a proposal with a regional approach, beyond the legislation of any specific country or administrative unit. ILDA started a second stage of this project that is currently supported by the Inter-American Development Bank (IDB) with an updated version of the standard in Honduras, Panama, Jamaica and, more recently, Ecuador.

- Strengthening cyber policy research centers.

An initial mapping developed by Derechos Digitales' Cyberpolicy Centre (IDRC, 2019) has identified 20 projects that will use automated decision-making systems to — allegedly — enhance State efficiency in responding to challenges in areas such as health, childcare, education, public security, access to justice, immigration, labour and transportation in eight different Latin-American countries.

The mapping project leaders noted that several of the initiatives target particularly vulnerable groups, exposing them to interventions to which they cannot give informed consent. In the subsequent stages of the project, they are aiming to develop in-depth case studies in different countries of the region to understand how automated decision-making systems are being implemented by states in the public sector (usually through public-private partnerships) and their possible impacts in terms of discrimination and violation of fundamental rights.

In a tweet

"Standardization of data is a very first step that helps organizations think and rethink how they are gathering, storing, accessing and publishing data." **- Silvana Fumega**

"All efforts to use AI for inclusion or social benefit must respect fundamental rights and never result in new forms of discrimination." **- Jamila Venturini**

5. Economy and Innovation

AI is poised to disrupt the economy in several fields. However, the extent and means of this disruption, and whether it will deliver a more inclusive economy, remains uncertain. The group agreed that AI research on innovation is not often linked with critical economic issues. Many countries in the Global South are still dealing with the consequences of the industrial revolution, which produced a massive divide in terms of wealth and opportunities, and some countries in the Global South are amongst the most unequal in the world. Having this conversation allows us to discuss how these countries can prepare themselves for what is being called a fourth industrial revolution. The group agreed that this discussion should include issues of taxation, automation of work and the development of local capacities and integrate them into governance debates. The group also agreed that AI technologies are often clustered in the hands of a few top players (mainly in the Global North) and it is unclear how the deployment of these technologies will affect the Global South. Previous research has found that the use of algorithmic techniques in labour has positive effects at the individual level. However, a structural discussion is necessary and Graham et al. (2019, p.289) have identified some frictions —imperfect information and alienation, discrimination, and the liability of foreignness— that need to be addressed when exploring how digital labour might fail to advance economic development goals. The group agreed that extensive use of these technologies could inaugurate a new era of capitalism where marginal costs will approach zero in many industries, where a "race-to-the-bottom" could occur in terms of wages and labor conditions in some industries, and several sectors could be radically reshaped. A critical asset of this unfolding revolution is data, and several Global North firms are hoarding significant amounts of it while those who produced the data — mostly data subjects — have little agency or rights. This has economic and ethical implications.

The group agreed that it is possible to have constructive debates with transnational corporations, although they are often not in attendance at forums where these topics are explicitly addressed, and this is an area of opportunity for future work on this subject.

The Web and the availability of data online provides several opportunities for firms to analyze data everywhere. However, increased datification does not necessarily mean that all data should, or could, be online and shared all the time, particularly in environments where connectivity remains an important issue. For example, the group agreed that it is worth exploring alternative data arrangements for particular groups, such as indigenous people or women, and such arrangements could lead AI to evolve differently in the Global South. The group agreed that we are still years away from a situation in which AI could effectively perform all human activities, and humans are very much needed at present to train, evaluate and decide when to use or stop using these systems. As local knowledge pours into algorithms designed by businesses with little understanding of context, the key question is whether this process is desirable, manageable and how it is going to serve — or not — the development of Global South countries.

Key questions

- How can we incentivize private investment to create vibrant ecosystems involving society, governments, academia and businesses? What are the limitations of this approach?
- What would alternative models of community self-governance of data look like?
- What are the consequences of having AI technologies perform tasks that require human involvement?

Cases

- CIPPEC survey on AI-powered technologies.

CIPPEC conducted a survey (Albrieu et al., 2019) of 307 companies from six branches of the Argentine manufacturing industry, in which they asked about the current and expected degree of penetration of new technologies and about the current and expected impact on the demand for work in each sector. The industrial sectors represented by survey participants included processed foods, steel and metalworking, light vehicles, textiles, farm machinery and biopharma. Some of the findings of this survey show that Argentina is just starting its journey into what some term "Industry 4.0." However, the survey found some similarities in the behaviour of certain groups of companies and classified firms according to their readiness to integrate new technologies into their business practice. 6% of the companies were classified as having the highest level of technological adaptation, 45% were classified as employing some technologies but still have room to achieve greater integration, and the rest (49%) were classified as having little to no integration of AI technologies in their business.

In a tweet

"We have to think about alternative models for disenfranchised communities to self-govern their data and infrastructure to ensure that they are able to exercise their rights to decide what to do with their data". **- Pyrou Chung**

"Societies must carefully consider whether AI should influence public discourse, even when trying to reduce hate speech and misinformation". **- Christopher Tuckwood**

"Economic tasks are being rearranged in order to create data-rich environments for machines to train and learn. In doing so, AI is creating a new division of labor between humans and machines". **- Ramiro Albrieu**

"How can we incentivize private investment to create these ecosystems where government, businesses, society and academia work together?". **- Priscila Chaves**

---

13. Results from the survey can be found at https://www.cippec.org/publicacion/travesia-4-0-hacia-la-transformacion-industrial-argentina/.

6. Governance

In this session, the group discussed key questions about data governance. Data governance is a relatively common term that has been discussed since the early days of the information revolution. Nevertheless, there seems to be less understanding of the importance of this issue in an environment in which only a few players are able to access, use and analyze large troves of different types of data with little oversight in an interconnected world. Determining who uses which data, for what purposes, for how long and what kind of liability they face are central challenges that do not have a common global answer.

The debate can seem sterile but it has actual implications. For example, Facebook, a global platform, was able to develop a census of African population density. The authors (Bonafilia et al., 2019) argue that this information could be used by governments, agencies, businesses and citizens alike to inform better policies and Africans to better understand their own contexts. In some cases, the Facebook census provides much more detailed and current information than official records. Nevertheless, there are currently no efficient and fair ways in which data collection, data use and data sharing can be applied, and, to date, most of these exercises are experimental. There is a very complex debate about the role large companies play, the value they add and the power they have vis-a-vis governments and societies.

The groups agreed there is a divide between Global North and Global South countries in terms of capacities and access to data. Ironically, users in the Global North could have more access to data from the Global South countries than the original producers of the data. By not facilitating the sharing and capacity-building at a local level, the data divide between data-rich societies and data-poor societies continues and in some cases, increases. However it is important to note that sharing data—even for specific purposes—also has risks, particularly in societies (both in the Global North and in the Global South) where data can be used to discriminate, prosecute or exploit vulnerable groups. The former is the reason to address data rights from an individual and collective perspective, amplifying the voices and concerns of specific

groups, empower their representatives and representation and eventually acknowledge their ability to opt out of these data regimes if they wish to do so. Governments need to step up and provide resources that contribute to capacity-building regarding data governance, training, management and awareness of value of data. Additionally, an accurate legal understanding is needed to assist government officials in the proper use of data.

As expected, privacy is a central concern in this line of thinking. The current European General Data Protection Regulation (GDPR) is considered the bare minimum needed to protect citizens' rights, but it does not have universal consensus. Furthermore, not all countries have the capacity or the will to implement this type of regulation as in other fields, countries may vary in how they will tackle privacy and data protection concerns, but this should not mean that central considerations about human dignity and human rights get left behind. Data protection and security should not be understood as a hindrance to policy makers, regulators and business. Rather, they should be understood as a way to foster trust and collaboration with data subjects. In short, the solution might not be a uniform global regulation but different types of regulation that vary according to context, respecting fundamental rights.

Finally, on the use of data, there is the issue of AI. AI techniques, most notably machine learning, offer a myriad of potential uses to automate, analyze and deploy actionable information. Experiments in the public sector abound and evidence about how it works remain scarce. With the availability of more timely, adequate, and higher-quality data, algorithms will be trained more efficiently, though not necessarily more effectively. Humans might still be needed to create, tune, supervise and address unexpected effects of automation in the public sector. The same applies in terms of economic and societal implications. Understanding how this would work in practice poses a great challenge.

Given the particularities and differences among societies, the group did not believe there needs to be a global unique answer to the several questions the issue of governance introduces. Not all aspects of human lives and societies need to be visible, datified and shared all the time, and not all records should live forever. Part of the group discussion focused on acknowledging that common frameworks are useful for thinking about these issues and their consequences, particularly when discussing whether data and AI can or should be used for a "good" agenda that can move society forward. Nonetheless, "good" is contextually defined, and what is good in one society might not be in another one. There are competing values and it is important to consider the legitimacy of AI use in a given context.

Key questions

- What is the current state of the art in terms of data governance for development? Can we move beyond "data as the new oil?"[14]
- Who and how should we discuss data governance arrangements and use of AI at a national and international level to ensure better development outcomes?
- How do we include vulnerable communities in the governance decision-making process to prevent exploitation and unintended outcomes?
- What aspects of gender should be considered in terms of data governance arrangements?
- How do we embed common principles respecting human dignity in data governance and AI for development?
- How do we ensure local appropriation of knowledge creation based on local data resources?

Cases

- Datos 360

Datos 360 is an initiative created by the National Agency for e-Government and Information Society (AGESIC, in Spanish) in Uruguay to promote a multidisciplinary approach to data management in Public Administration. The objective of Datos 360 (AGESIC) is to incorporate different perspectives of the treatment of public data, including Personal Data Protection, Access to Public Information and Data Security, among others. Some results of this strategy have been a Data Policy that established a national strategy that is data-driven and established the design, delivery and monitoring of public policies and services. Other results of this initiative are the data governance and data architecture guidelines and an Artificial Intelligence strategy that promotes the responsible use of AI in public administration.

- Privacy International's Examples of Artificial Intelligence related harm

Privacy International mentioned that their organization maintains a database of examples of abuse of data. In this database, it is possible to filter by type of technology, such as abuses that are powered by AI. Some of the examples mentioned on their website that are related to AI are related to the uses of AI in gambling and the risks this poses for young adults, increased use of automated decision making in immigration decisions in the United States and Canada and the effects of predictive policing. Other areas of AI use where harms were detected are related to automated translation, facial recognition and other technologies commonly used in social media. A common theme throughout the examples is that countries' regulatory frameworks are not yet adept at regulating the use of these technologies and their possible uses for harm, presenting a challenge for both the Global North and the Global South.

In a tweet

"Governments seeking to regulate data have to go beyond what they see as personal data and beyond the logic of control of data. If disclosure, transparency and notice are not enough, what form should regulaton take?" **- Christian Perrone**

"It is time to reimagine citizenship in the digital world and go beyond this idea of digital personality as the only way to be a citizen" **- Grace Mutungu**

"We work with AI in the government and what we want is to give guarantees to the citizens that the public organizations are going to be making good use of AI in public administration"
- Laura Rodriguez

---

14. For more information on the concept of "data as the new oil," we recommend looking at Clive Humby's blogpost on the subject at https://akasha.org/blog/2019/01/21/interpersonal-data-2-of-3

7. Capacity building

Several of the groups represented in the workshop are working on different facets of the intersection between AI, human rights and development, including on privacy, openness, artificial intelligence, and right to information, among others. However, many of the communities that work at this intersection do not always communicate with each other, which prevents them from sharing the knowledge and capacity they have built. Moreover, there are cultural aspects to consider when thinking about knowledge sharing and the application of AI technologies, such as gender and broader inclusion. Many organizations are not yet prepared to address these cultural aspects and could benefit from integrating them into their knowledge-sharing strategies. Overall, workshop participants discussed how quality data and talent are essential for AI benefits to fully flourish in the Global South.

The group agreed on the need to advance a core standard regarding data capacities across several fields. This would evolve in different curricula aimed at specific groups, but based on the same skills and with potential applicability across Global South countries. For instance, middle management across the public sector, NGOs and business organizations could benefit from acquiring a basic understanding of data science. In addition, public oriented specialists could benefit from acquiring a basic understanding of how AI operates and they would be able to develop a critical outlook about when and how AI could deliver good value to the public.

Formal and informal education systems have a key role in this area. Open education approaches and certifications could create more opportunities for the Global South. Furthermore, corporate initiatives could participate in these efforts by sharing capacity and knowledge and by creating local value.

Key questions

- How do we best integrate middle management public administrators in capacity building strategies?
- In what ways can the demand for more data scientists be addressed in the Global South context?

Cases

- Khipu conference in Montevideo

The Khipu conference (Khipu, 2019) was a meeting of Latin American Artificial Intelligence professionals that took place in Montevideo from November 11 to 15, 2019. Khipu consisted of five days of advanced training and talks with keynote speakers such as Jeff Dean (Director of AI at Google) and Ian Goodfellow (Director of Machine Learning at Apple). This conference was inspired by the Deep Learning Indaba[15] conference that is held in various African countries and was organized by Latin American researchers with a combined conference and summer school model.

One of the main reasons for organizing these conferences was that the original organizers of Indaba noticed that there was hardly any representation of African and Latin American researchers in the discipline's main conferences, and they believed that creating their own meeting was the way to bring talent to the region.

- The Carribbean Open School of Data Blended-Learning Model

The Carribbean Open School of Data (COSD) implements a blended-learning model (COSD, 2020) which comprises three main areas: learning analytics, facilitated sessions and self-paced e-learning. The key principles behind this model are research-based needs analysis and learner profiling, using a "flip the classroom" model, facilitating access to content that is mobile-enabled and device-agnostic, and facilitating access to experts and context-relevant reference materials.

Specifically, a project in Haiti where this blended learning model was applied was mentioned because they combined long-distance training with on-the-ground training, and this is particularly useful in low-tech contexts, where capacity building is most needed.

- Data literacy training in the Mekong region

The Open Development Initiative has provided data literacy training in Myanmar, Cambodia and Thailand, specifically to build the capacities of civil society, journalists, public servants, think tanks and other stakeholders to work with data. The modules developed were intended to promote a data-driven culture, and to realize the positive benefits of investment in data training and upskilling. Each participant is chosen specifically to ensure that skills are retained within institutions where data is in demand but as yet under-utilized. The intensive 3-week course is delivered over several months, and covers skills encompassing the entire data value chain from collecting, cleaning to final publication. Each participant is expected to produce a data-driven end product that can be utilized in their work setting.

Although the focus is upon upskilling data literacy, a significant aspect of the training includes data protection, security and ethics. All are necessary topics in the Mekong context because data is political and there are currently no adequate safeguards in place that protect citizens from constraints on freedom of expression put in place by oppressive regimes.

---

15. Read more in https://deeplearningindaba.com/2020/.

8. Scenarios

A scenario methodology can be used to discuss "detailed descriptions or stories of plausible future events and outcomes" (Wilson 2003 in Tighe 2019, p.16.) as well as the "evolving dynamics of interacting forces rather than the static picture of a single end-point future" (Wilson, 2003 in Tighe, 2019, p. 16). In the context of the workshop, the scenario methodology was considered useful because we aimed to generate scenarios to explore complex future contexts in which organizations would need to operate. Exploring alternative scenarios allows organizations to be resilient in the face of change, make decisions and reframe interventions, and to adjust to or prevent certain imagined futures.

According to Wilkinson and Kupers (2014), scenario planning was initially developed to allow for a more adaptive approach than forecast-based planning, making uncertainty part of the strategy and incorporating the need

for flexibility in planning by focusing on what might happen that is beyond the control of the organization. A scenario building toolkit (Noortman, Koning, Vervoort & Hoofd, 2019) on which we based our workshop component emphasizes that the usefulness of the technique does not depend on the likelihood of the scenarios coming true, but, rather, on the development of a flexible response.

Due to the time constraints of the workshop, we concentrated on the first part of the methodology, spending time discussing future possible scenarios amongst different groups of participants. The backgrounds of the participants, their expertise and experience provided an opportunity to start exploring the future of algorithm-based decisions in the public sector along with their potentials and risks. In the workshop, participants were divided into subgroups to discuss how the future might look if a specific stakeholder were the one "calling the shots."

The questions discussed included the following:

- *What would be the risks?*
- *What would the future look like?*
- *What are the drivers of change?*
- *Who would win or lose in a specific scenario?*
- *What kind of policies need to be in place?*
- *What kind of research do we need to invest in?*
- *What are the economic, political, environmental, technological and societal factors in each scenario?*

In the next section, we present the reflections of each of the groups, as well as a set of follow-up questions we believe need further examination. Pursuing these follow-up questions would allow us to develop in future work the flexible responses that are part of the scenario methodology.
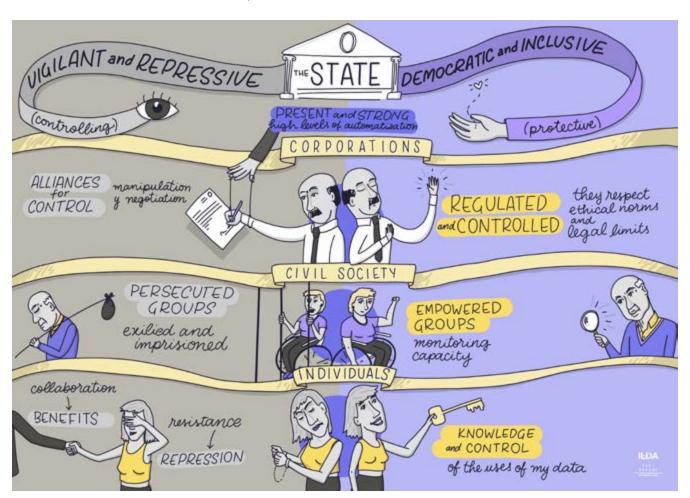
In this section, four scenarios were discussed in teams to which participants were randomly assigned. The presentation of the following section is structured according to the scenarios:

1. Presence of a strong state

In the future, the State is present and is able to dominate the data revolution. The State develops abilities to access, regulate, control and use data in a society. The State has analytical capacity and uses data in unprecedented ways, which helps guide their objectives and society in general. In this scenario, two different primary trajectories were identified:

I.   The state uses these new capacities and powers to regulate and control the abuses of the private sector regarding citizens' personal data and also establishes frameworks to ensure that sharing data benefits the common good. Society serves as a counterweight to the power of the State.

II.   The State uses these new capacities and powers to regulate with unchecked power, to achieve social control. With total control over the data or in alliance with the private sector, the State becomes an omnipresent actor that regulates people's lives, using the data to fulfill its objective. In this scenario, the State is totalitarian and individual rights are not guaranteed or protected.

Participants also discussed the fact that this scenario, particularly in its first iteration, might result in an enhanced and dignified professional civil service that is oriented towards the needs of citizens and that would pursue evidence-based public policy strategies. Red flags indicating that the State is proceeding along the second, totalitarian pathway could include the excessive use of surveillance measures, the state perpetrating violence with impunity, and centralization of power.
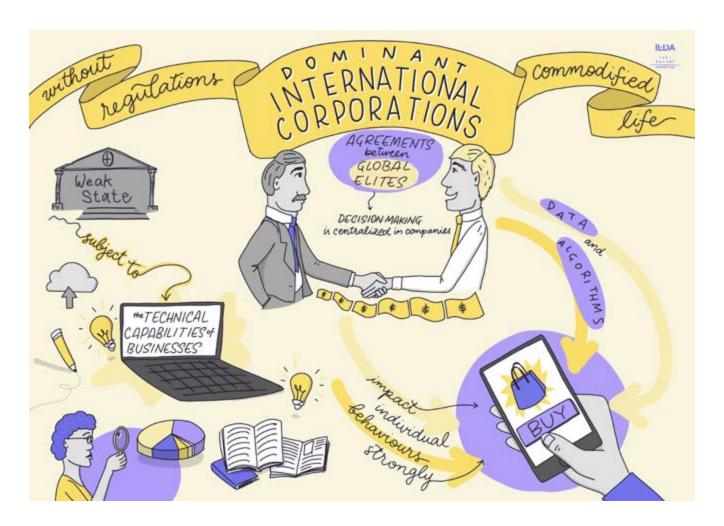
2. Multinational corporation domination

In the future, a group of dominant corporations, based primarily in the Global North, have control over people's data. States depend on the corporations to fulfill their functions, and corporations have taken over decision-making organs in an integrated and globalized world. The use of data and algorithms is widespread in society, where it is linked to the sale of goods and services of all kinds. The sustainability of the infrastructure needed in this scenario comes from agreements between the dominant organizations, that are governed mainly by economic principles.
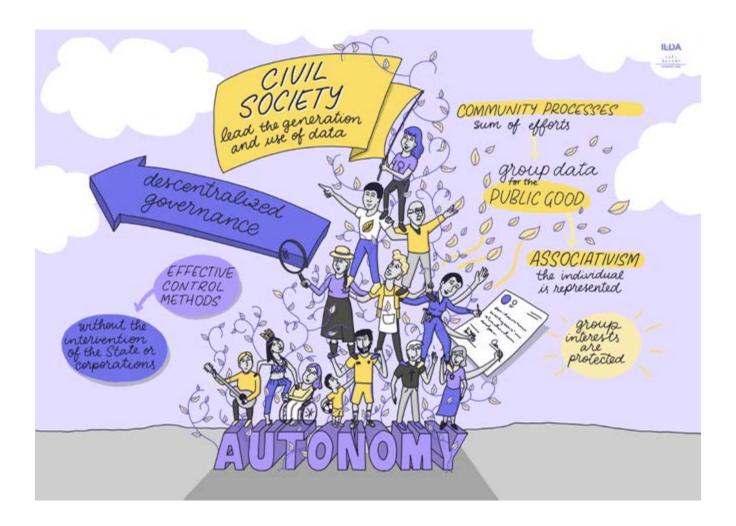
Drivers of change:

- Piles of money;
- Complacency on the part of users;
- Talent + innovation;
- Digitalization of everything.

Participants also discussed the fact that key actors in this scenario would be corporations from the United States and China, mainly telecommunication corporations, data brokers and financial institutions.

3. Civil society leads the production and use of data

In the future, civil society organizations, local governments and some states are able to generate an autonomous space to use their data to pursue community goals. Governance is decentralized, allowing specific groups (ie. indigenous groups, communities) to govern the use of their data and have control over the technology that they use. There are imbalances in rural and urban contexts. In this context, people have a high degree of autonomy and liberty and there exists a balance between the power of the public and that of the private sector that provides services. Participants also discussed the fact that some challenges in this scenario would be the duplication of initiatives and an increase in competition for resources to fund organized community efforts.

4. My data and my decisions

In the future, people can determine how their data will be used. From the genetic code to their spatial movements, people can donate or sell their data and have total control over them. There will be agile markets that allow for the commercialization of data and there will be distinctive economic, social or cultural benefits for doing so. People will have different ways in which to opt out of these agreements, which, in general, will be with corporations or businesses. In this future, the focus is on the individual, although there is a possibility that groups of people can coordinate collectively to negotiate over the use of their data. There is interoperability and smart contracts that make this process frictionless, and it generates new forms of socialization and business opportunities.

Participants also discussed the fact that centralized governments could disappear in this scenario because it is hard for authorities to enforce control over people who control their own data assets. However, other actors — mainly corporations — may arise as centralized forces, as they are the ones with whom the data is traded.

9. The road ahead for data governance and AI: A new fair data deal?

The current state of the art with respect to the use of data and AI for development is complex, and the opportunities, risks and potential threats of the use of data and AI for development outcomes is not yet properly understood. In the coming years, markets, governments and citizens could shape the evolution of their societies by using these tools to provide better and more inclusive futures or to promote and exacerbate existing tensions in our interconnected world. From the discussion, we identify a set of initial ideas that should evolve into research and policy conversations in their own right.

9.1 Connecting the dots: Data for the Public Interest[16]

Public, private and non-governmental organizations and individuals working on open data, big data, responsible data and AI for good are dealing with the same problems although from different viewpoints and with different stakeholders. There is a larger picture emerging in terms of how to use data for the public interest, in very distinct contexts. Basic questions about who, when, how, to what degree and to what end have to be considered when planning to use data to foster development in different contexts — such as economic sectors and communities — in which this data will be used.

In the same way, whether automation or, more generally, the use of AI can deliver a more inclusive, fair and sustainable society remains to be seen. There is a need to foster a dialogue between communities having different perspectives and between disciplines to understand how what seems to be an unstoppable trend could evolve in the coming years. There is also a need to connect data and AI in

---

16. "Public interest" here is defined in a way that includes a multi-stakeholder view of interest and not just a government or state view.

the same dialogue, as machine learning techniques can only be as good as the data used to train them. Data cannot be seen just as a resource for decisions to be made, it has to be understood as a representation of relationships, people, societies, productive systems and, as such, is very different from an extractive natural resource such as oil, minerals, or timber, despite common comparisons.

9.2 More diverse voices from the South

To date, these debates have been dominated by countries and companies based in the Global North. Participants in these debates come with certain biases in terms of how platforms, tools and data actually work and how this impacts other societies. After more than 20 years of internet expansion and digitalization, there is enough evidence to consider the importance of facilitating greater participation of Global South communities in policy and technical developments. Moreover, there is also a need to include marginalized communities who are at the receiving end of these developments. Without the active participation of groups who are often excluded from the design and implementation of technological solutions, these solutions will perpetuate their exclusion, potentially jeopardizing their rights and ways of life. Acknowledging this reality can lead to the development of governance structures around key technological infrastructures that reflect the diversity and power imbalances across the world.

9.3 Enabling a global debate for a fair data deal

Data is now central for all human endeavours but questions about justice, equity and ownership seem absent from global forums. There have been some attempts in international forums where AI, digital development, and data for development are discussed, and these discussions offer valuable insights into the relevant topics and debates regarding these issues. Nevertheless, participants in these discussions are often not representative of voices in the Global South, nor are they very diverse.

Furthermore, the analytical framework used in some of these debates does not explicitly address the need for an inclusive or fair data deal which links the technical aspects of data with its social and normative implications in terms of human rights, use for development and collective ownership of data. If data is to be used to address several development challenges, a framework that understands and promotes ethical and lawful uses of data and strategies to mitigate risks is essential. Such frameworks require technical expertise, but ultimately decisions require political will from both societies and governments to implement them successfully. At the moment, such decisions are often ill-informed and

are made on the basis of need and urgency, often producing undesirable consequences.

9.4 Data in context

Data means different things to different audiences, and is used in different ways. For instance, having data about femicides helps reveal a social phenomenon that is not often understood or acknowledged in our societies. But publishing data about violence in certain territories or contexts without adequate planning — through established privacy measures or educational campagins — could lead to discrimination against entire populations. Data by themselves are not always the problem; instead, the use and framing that entities give to the data is generally where issues develop. There is a need to develop sectoral approaches to data governance and AI to explore, regulate and use data effectively. Initiatives around agriculture, health or transparency are good examples of how specific communities can advance understanding of and policy concerning data governance and the use of machine learning techniques. To advance this agenda in a direction that includes ethical, human rights and context considerations, we should support the creation of diverse task forces of experts to provide analysis and recommendations in each field or for each relevant project.

9.5 Imagining and shaping a different digital future

Since the beginning of the digital age and the Web, the world has faced uncertainty about how these tools will impact the way the global community organizes itself. Utopian and dystopian scenarios are possible, but an approach grounded in current trends is often neglected. William Gibson's famous quote "The future is already here – it's just not very evenly distributed" is often inaccurate. We are still able to shape the future, as the world deals with different experiences in terms of digitalisation, data governance and use of AI techniques that could lead to very different futures.

Moreover, an active approach to monitoring current trends based on evidence could inform more refined scenarios about how different societies could evolve, and what kind of coordination is needed for the world to address common challenges such as health issues, resource management or climate change. Imagining the future is not a task exclusively for futurists or enlightened individuals, it is a task that should be democratized and one in which organizations and individuals with different experiences and contexts should participate. As the first wave of digital disruption comes to an end, having concentrated more wealth and power in fewer hands, with local communities having little say about their future, it may be time to reimagine how to
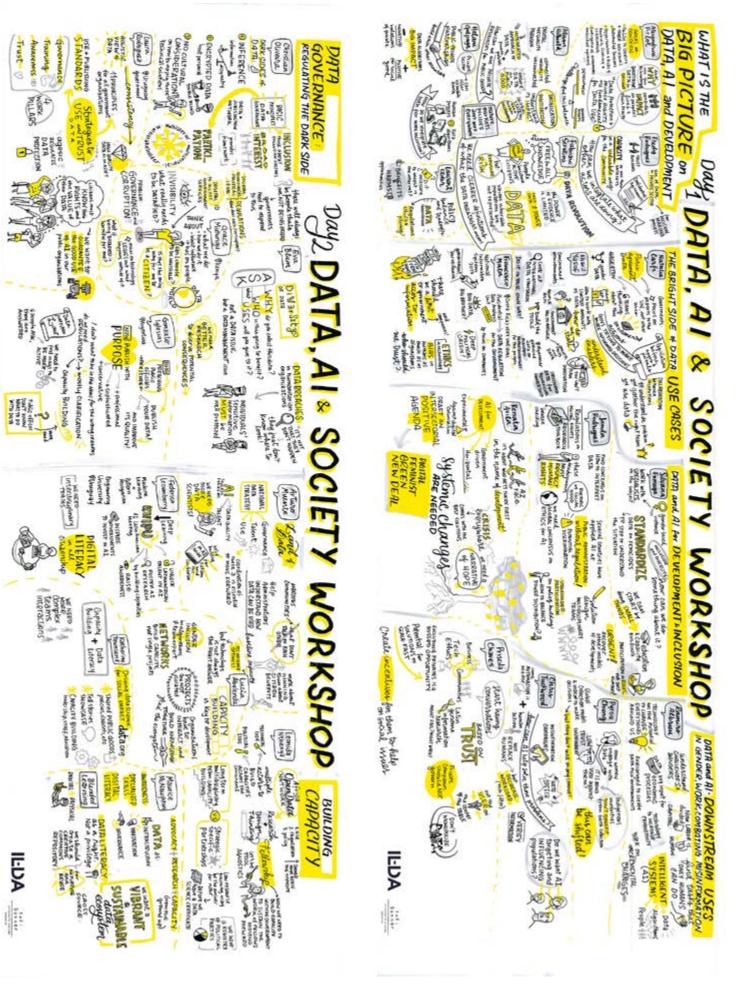
shift the current state of affairs towards broader ownership and greater accountability and inclusion.

9.6 Addressing the capacity gap

AI and data talent are not evenly distributed, and this uneven distribution is not exclusive to the Global South; the Global North has many of its own issues regarding unequal access to these technologies. By "data talent" we mean the capacity to build an inclusive, safe and fair ecosystem. The talent needed is not only on technical skills but also on the right way to manage, regulate, evaluate and adapt technological developments. These are the so-called "soft" skills which are crucial for governing the development of any technology. In both contexts, the structures that permeate unequal access and distribution are complex. However, it is necessary to acknowledge that, due to historical patterns of uneven accumulation, the Global South faces more constraints than the Global North in terms of access to resources to combat these inequalities.

Furthermore, data and AI-related work require different sets of skills. For example, tagging data for algorithms to learn is already happening in the Global South, providing some value to workers, but this work does not enhance digitalisation, data governance or use of AI techniques in ways that could lead to very different futures.

We can still shape a more inclusive, fair and open use of data for development. To do so we need dialogue, imagination and acknowledgement of the complexities this agenda will bring in the years to come.

References

1. AGESIC (2019) *Datos 360: Visión integral del uso de datos del estado*. Available at: https://www.gub.uy/agencia-gobierno-electronico-sociedad-informacion-conocimiento/comunicacion/noticias/datos-360deg-vision-integral-del-uso-datos-estado-0 (Accessed: 21 May, 2020)

2. Albrieu, R., Basco A.I., Brest López, C., de Azevedo, B., Peirano, M and Vienni, G. (2019) *Travesía 4: hacia la transformación industrial argentina*. Available at: https://www.cippec.org/wp-content/uploads/2019/06/Traves%C3%ADa-4.0-hacia-la-transformaci%C3%B3n-industrial-argentina.pdf (Accessed: 21 May, 2020)

3. Ananny, M., & Crawford, K. (2018). Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Society, 20*(3), 973–989. https://doi.org/10.1177/1461444816676645

4. Bonafilia, D.,Gill,J., Kirsanov.S, & Sundram,D. (2019) *Mapping the world to help aid workers, with weakly, semi-supervised learning*. Available at: https://ai.facebook.com/blog/mapping-the-world-to-help-aid-workers-with-weakly-semi-supervised-learning.(Accessed: 21 May, 2020)

5. Canares,M., Yusof, K., Men, S. (2017) *Collaborating for open data. Building an open database on politically exposed persons in Malaysia: A case study*. Washington DC: World Wide Web Foundation

6. CSOD (2020) *Blended learning model* Available at: https://drive.google.com/file/d/149re-KD2E98KcaLMsSIN_89HvrJCMK-G/view (Accessed: 21 May, 2020)

7. Davies, T., Perini, F., & Alonso, J.M. (2013). *Researching the emerging impacts of open data – ODDC conceptual framework*. Working Paper. Washington, DC: World Wide Web Foundation. Available at: http://www.opendataresearch.org/sites/default/files/posts/Researching%20the%20emerging%20impacts%20of%20open%20data.pdf .(Accessed: 8 July, 2020)

8. Davies, Walker, Rubinstein, and Perini (eds.) (2019) *The state of open data: histories and horizons*. African Minds, International Development Research Centre.

9. Eaves, D. (2012, September 7) Lies, *Damned Lies, and Open Data. Slate*. Available at: https://slate.com/technology/2012/09/open-data-movement-how-to-keep-information-from-being-politicized.html (Accessed: 8 July, 2020)

10. Facebook (2019) *High resolution population density maps*. Available at https://data.humdata.org/dataset/highresolutionpopulationdensitymaps.(Accessed: 21 May, 2020)

11. Fumega, S. (2019) Open data, gender and violence in Latin America. *European Public Mosaic* (7). Available at: http://www.gencat.cat/eapc/epum/N7/pdf/EPuM7Fumega.pdf (Accessed: 8 July, 2020)

12. Graham, M; Hjorth, I, and Lehdonvirta, V. (2019) Digital labor and development: Impacts of global digital labor platforms and the gig economy on worker livelihoods in Graham, M. (ed) *Digital economies at global margins*. MIT Press. pp. 269-294.

13. IDRC (2019) *Strengthening cyber policy research centres in the Global South – Derechos Digitales*. Available at: https://www.idrc.ca/en/project/strengthening-cyber-policy-research-centres-global-south-derechos-digitales (Accessed: 21 May, 2020)

14. Jütting, J. and McDonnell, I. (2017) *Overview: What will it take for data to enable development?* OECD. Available at: https://www-oecd-ilibrary-org.ezproxy.library.ubc.ca/docserver/dcr-2017-6-en.pdf?expires=1591293289&id=id&accname=guest&checksum=BB9CC3D-249D243610DAECE3A25BCA4F6 (Accessed: 4 June 2020)

15. Khipu (2019) *¿Qué es Khipu?* Available at: https://www.fing.edu.uy/sites/default/files/eventos/khipu-inteligencia-artificial/Khipu%20PR.pdf (Accessed: 21 May, 2020)

16. Noortman, Koning, Vervoort and Hoofd (2019) *Imaginative scenario planning toolkit*. Centre for Research and Evidence on Security Threats. Available at: https://crestresearch.ac.uk/resources/imaginative-scenario-planning-toolkit/ (Accessed: 21 May, 2020)

17. O'Donnell, M., Calderón, A and Strickland, C (Draft) *Closing Gender Pay Gaps: Identifying Roles for Government & the Private Sector*. Available at: https://docs.google.com/document/d/1cfzwwAWKPfDdekrnmntf9twgig_fxQVPLCZLBatJMyw (Accessed: 8 July, 2020)

18. Privacy International (2019) *Examples of abuse database*. Available at: https://privacyinternational.org/examples?field_type_of_abuse_target_id_2%5B%5D=264 (Accessed: 21 May, 2020)

19. Smith and Neupane (2018) *Artificial intelligence and human development: toward a research agenda*. Available at: https://idl-bnc-idrc.dspacedirect.org/handle/10625/56949.(Accessed: 21 May, 2020)

20. Taylor, L. (2016). The ethics of big data as a public good: which public? Whose good?. In *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2083).

21. The Sentinel Project (2020) The role of technology. Available at: https://thesentinelproject.org/what-we-do/the-role-of-technology/ (Accessed: 21 May, 2020)

22. The Sinar Project (2019) *Parliamentary answers as source for data demand and government oversight*. Available at: https://sinarproject.org/open-parliament/updates/parliamentary-answers-as-source-for-data-demand-and-government-oversight. (Accessed: 21 May, 2020)

23. Tighe, S. (2019) *Rethinking strategy. How to anticipate the future, slow down change and improve decision making*. John Wiley & Sons

24. United Nations (2015) *A world that counts*. Available at: https://www.undatarevolution.org/wp-content/uploads/2014/11/A-World-That-Counts.pdf(Accessed: 21 May, 2020)

25. Van Schalkwyk, F., Verhulst, S.G., Magalhães, G. and Pane, J. (2017) *The social dynamics of open data*. African Minds. Available at: https://www.africanminds.co.za/the-social-dynamics-of-open-data/ (Accessed: 8 July, 2020)

26. Weaver, C., Powell, J., & Leson, H. (2019) "Open data, development assistance, and humanitarian Action". In T. Davies, S. Walker, M. Rubinstein, & F. Perini (Eds.), *The state of open data: histories and horizons*. Cape Town and Ottawa: African Minds and International Development Research Centre. Available at: https://www.stateofopendata.od4d.net/chapters/sectors/development-humanitarian.html .(Accessed: 25 May, 2020)

27. Wilkinson, A., and Kupers, R. (2014). *The essence of scenarios: learning from the Shell experience*. Amsterdam University Press.

28. World Bank Group (2020) *World development report*. Available at https://www.worldbank.org/en/publication/wdr2020.(Accessed: 21 May, 2020)

29. World Wide Web Foundation (2018) *Open data barometer*. Available at: https://opendatabarometer.org/doc/leadersEdition/ODB-leadersEdition-Report.pdf.(Accessed: 21 May, 2020)