United Nations | Department of Economic and Social Affairs

DECADE OF ACTION

POLICY BRIEF NO89

Strengthening Data Governance for Effective Use of Open Data and Big Data Analytics for Combating COVID-19

BACKGROUND & CONTEXT

The fast spread and pervasive impact of COVID-19 require governments to provide effective, timely, and inclusive responses to manage the pandemic. In addition to traditional data sources and data analytics tools, governments rely on *open data* and *big data analytics* in responding to COVID-19. Open data and big data analytics are required for i) conducting real-time situation analysis, contact tracing, and early and timely diagnosis for effective containment; ii) facilitating coordination and collaboration between national and local governments and fostering the ownership and accountability of local governments; iii) securing public trust in government through better transparency and improved communications; iv) countering misinformation; v) identifying and addressing special vulnerabilities and needs of vulnerable groups by gathering disaggregated data; and vi) supporting effective management of medical equipment supplies and demands.

"Open data" are data which are made accessible and available in a standardized machine-readable format and under a license that allows it to be re-used and re-shared. Because the type, format and quality of data vary significantly, it is a challenge to certify the data and put them in effective use with value added. "Big data" are usually associated with high velocity, volume and variety and often defined within political and social contexts as "a cluster or assemblage of data-related ideas, resources, and practices".[1] Big data are also referred to as an "imprecise description of a rich and complicated set of characteristics, practices, techniques, ethical issues and outcomes all associated with data".[2] There are many open and big data sources such as open government data, citizen science, and crowdsourcing that provide updated information. Big data analytics can be used for deeper and more complex tasks such as the analysis of social media sentiment. Particularly, citizen science, which refers to the involvement of non-scientist citizens in the generation of new scientific knowledge, could effectively contribute to building resilient communities.[3]

With a more prominent role of open data and big

Summary

Governments are highly dependent on all data including official statistics, administrative data, open data and big data analytics for decision-making and actions to address the COVID-19 pandemic. These data allow governments to set priorities and adjust their decisions quickly and effectively in response to rapidly evolving COVID-19 situations. Open data and big data analytics, particularly through Artificial Intelligence (AI) platforms and data visualization tools, are empowering governments to predict virus mutations, track virus spread in real-time, and identify medications for treating COVID-19. Governments are using big data analytics to get prepared, react effectively, and develop both short-term and long-term strategies. Yet, increasing public concerns about data privacy and security put in jeopardy public trust in data collection, use and dissemination by government, business and relevant non-government institutions. To ensure effective use of open data and big data analytics for combatting COVID-19, it is necessary to strengthen data governance with regard to data collection, data partnership, data analysis, data dissemination, and protection of data privacy and data security.

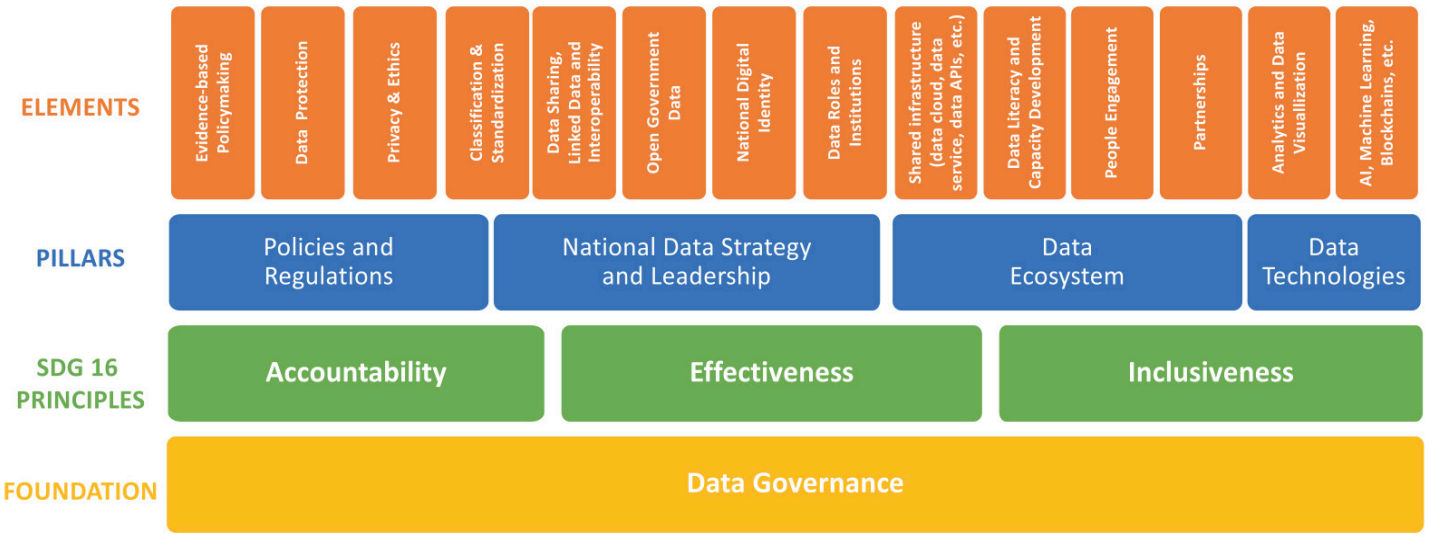
A holistic and whole-of-government approach to data governance will not only help address public concerns about data privacy and enhance public trust but also support institutional coordination, mobilize concerted actions, and streamline government operations as a whole.

1 UN DESA (2020). UN E-Government Survey 2020, p. 148
2 Ibid.
3 UN Economic and Social Council, The 22nd Session of the Commission on Science and Technology for Development (March 1999) E/CN.16/2019/3

Authors: Keping Yao and Mi Kyoung Park, UN Project Office on Governance, Division for Public Institutions and Digital Government, UN DESA
For further information, contact undesa@un.org, or visit www.un.org/development/desa/publications/

Figure 1

Illustrative data governance framework for e-government



Source: UN DESA (2020). UN E-Government Survey 2020, p. 166.

data analytics in addressing challenges of COVID-19, there have been increasing concerns about data privacy and security, which also put in jeopardy public trust in data collection, use, and dissemination by government and relevant non-government stakeholders such as the private sector. According to a recent survey by IBM Policy Lab, almost half (49%) of the respondents across the US and the EU are more concerned about privacy as a result of contact tracing during the COVID-19 pandemic.[4] This has become more complex due to the increase in data partnerships between government institutions and the private sector such as telecommunications and credit card companies for purposes of contact tracing and social distancing. The general public has expressed concerns about the unauthorized access and misuse of personal data as such access is often conducted without consent. The public is not duly informed about the legal or technical protocols governing such data partnerships between government institutions and the private sector. Such increasing public concerns about data privacy and security with regard to the use of open data and big data analytics for addressing COVID-19, which is often linked with data partnerships between government institutions and the private sector, have led to calls for strengthening data governance.

Data governance is a systemic and multi-dimensional approach to setting policies and regulations, establishing leadership for institutional coordination and national strategy, nurturing an enabling data ecosystem, and streamlining data management. An illustrated data governance framework for national e-government is shown in *figure 1*. As reflected in the four pillars shown in the figure, data governance is supported by the dynamic relationship between policies, institutions, people, processes, and enabling technologies.

Effective data governance is required to tackle the complexity of multiple and interdisciplinary data sources and the significant challenges of coordinating data gathering and collection, data partnerships, and data analytics. It is also crucial for facilitating data dissemination across government institutions and between government and citizens and other stakeholders. Data governance creates an enabling environment for sharing data, linking data, interoperability, and data exchange; it often includes initiatives in terms of data generation, reporting, result dissemination, and data innovation.

Establishing a framework for effective data governance is challenging due to a multitude of reasons. Particularly, in harnessing open data and big data analytics for responding to COVID-19, governments are faced with some long-standing challenges in data governance. These challenges include i) *lack of legislative and policy frameworks and technical means to protect personal privacy*; ii) *lack of sufficient policy, institutional and technical coordination on data partnership between government institutions and relevant stakeholders including the private sector*; iii) *insufficient measures to safeguard data security and ICT infrastructure*; iv) *inadequate and insufficient*

4 https://www.ibm.com/blogs/policy/morning-consult-pandemic-tech-privacy-poll/

data disaggregation, which leaves the vulnerable groups unaccounted for or left out from government rescue measures or emergency assistance; v) *lack of effective risk communication strategy*, which is essential for building public trust; and vi) *lack of a data ecosystem*, which causes data fragmentation and data silos.

In times of public health emergencies, these problems constitute severe barriers to generating public value of open data and big data analytics, which inevitably set back innovative solutions for effectively addressing emergency situations.

KEY ISSUES ON INSTITUTING DATA GOVERNANCE FOR ADDRESSING THE COVID-19 PANDEMIC

Protection of Personal Data Through Laws and Regulations and Policy Frameworks

To effectively track and contain the COVID-19 virus, governments may have to collect and gather a large pool of personal data such as mobile network data and credit card data, and public security surveillance data, which has raised public concerns of protection of personal data, especially regarding the authorized access to such data, secure data storage, specified data use, and anonymized individual data dissemination. Registered public concerns relate to the legitimacy of collecting and using personal data, the lack of public awareness, and the lack of clarity as to when and how the personal data would be used. There are also worries about the proliferation of personal data collected and the profiling and surveillance applications used by government and the private sector to gather information on the population. In many cases, government may not seek consent for public use, and citizens may lack information on the protocol governing the data partnerships between government institutions and the private sector.

Data ownership during such partnership is not always clear, especially when data management is shared or transferred between agencies, which makes it difficult to assign or trace accountability. A more serious situation is the possible stigma against certain individuals and communities or specific groups of people in society, if the data related to these individuals and specific groups of people become publicized without their consent or unlawfully. Moreover, there have been some incidents of conflict between private data ownership of certain groups, such as religious groups, and public data governance for the interests of society as a whole. The latter calls for a general consensus on clear legal grounds to enforce the submission of data to protect the lives of the majority.

It is, therefore, critical to enact personal information protection acts or revise the existing laws and regulations. For example, in the Republic of Korea, there were amendments to the 'Three data protection laws (The Personal Information Protection Act, The Information and Communication Network Act, and The Credit Information Act)' in 2020, which particularly brought together personal information protection functions scattered across different ministries before.[5] Specifically, the amendments to the Personal Information Protection Act include clarification of the definition of personal information, provisions on the processing and combination of pseudonymized data, expanded use of personal information within the scope reasonably related to the initial purpose of the collection and in consideration of the necessary measures taken, and strengthened safety obligation measures for a personal information controller.[6][7] Also, based on the amendments to the "Three data protection laws", the reformed Personal Information Protection Commission (PIPC) was launched as a central administrative agency under the Prime Minister with an integrated supervisory authority on personal information protection.[8] In Spain, there is a GDPR-related clarification report[9] on the usage of digital technologies. The Report notes that the Personal Data Protection Regulation, while aimed at safeguarding a fundamental right, is applied in its entirety during the COVID-19 pandemic in relation to the data processing resulting from the current situation and from the spread of the virus. The Report also clarifies that there is no reason to suspend fundamental rights with regard to data protection during the pandemic.

Many governments have articulated policies in an effort to promote protection of data privacy. The recent four editions of UN E-Government Survey indicate that the number of countries with privacy statements available online increased from 97 in 2014 to 128 in 2020 (*figure 2*).

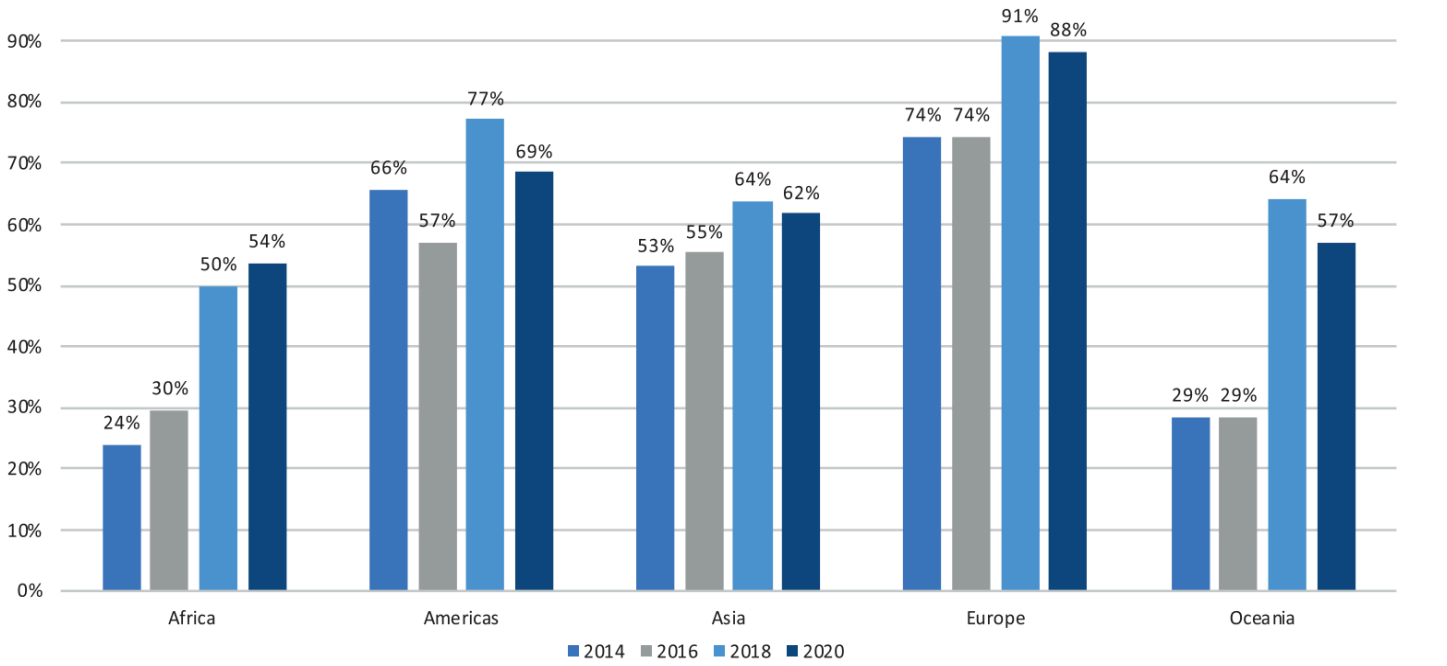
Mobilizing Whole-of-Society Efforts and Increasing Engagement of Stakeholders Through Policy, Institutional and Technical Coordination

A centralized data platform for coordinating data partnerships is critical for facilitating effective

5 https://www.pipc.go.kr/cmt/english/introduction/chairman.do
6 The Personal Information Protection Act. Available at: https://www.pipc.go.kr/cmt/english/news/selectBoardList.do?bbsId=BBSMSTR_000000000128
7 Personal Information Protection Commission (2020). 2020 Annual Report on Personal Information Protection. Available at: https://www.pipc.go.kr/np/cop/bbs/selectBoardList.do?bbsId=BS079&mCode=D070020000
8 http://www.pipc.go.kr/cmt/english/news/selectBoardArticle.do?nttId=6697&bbsId=BBSMSTR_000000000124
9 https://www.aepd.es/es/documento/2020-0017-en.pdf

Figure 2

Countries with privacy statements available online



Source: UN DESA (2020). UN E-Government Survey 2020, p. 162.

collaboration among different stakeholders and mobilizing whole-of-society efforts for an effective COVID-19 response. It assists with a comprehensive and timely understanding of needs of other stakeholders, such as the private sector, IT industry, expert groups, academia, civil society organizations, and citizens. It provides information on the areas where government is experiencing gaps in capacity, requiring contributions by various stakeholders. For example, the Japanese government established a public-private joint task force "Tech Team Responding to COVID-19" to promote the utilization of data for limiting the spread of COVID-19.[10] The taskforce shares experiences between organizations and deploys services while collaborating with the private sector.

The engagement of multiple stakeholders is particularly important for effective data collection, data analysis, and data partnership. The private sector and IT industry, for example, can provide technical and financial resources for effective data governance, such as providing data sharing platforms and sharing information for contact tracing. Expert groups and academia can assist governments with collecting, interpreting, and analyzing data, including medical data. For example, in Saudi Arabia, researchers at King Abdullah University of Science and Technology (KAUST) are coming together to understand SARS-CoV-2 to help protect communities. They are using the supercomputer of KAUST to not only compare and analyze COVID-19 cases but also to scan billions of environmental samples for traces of SARS-CoV-2. Their work is already available as an open research tool, providing access to all data to scientists around the world.[11] Civil society organizations can also support government in generating disaggregated data and disseminating data to vulnerable groups such as older persons, persons with disabilities, women, and youth. Furthermore, government can tap into the ideas, knowledge, and skills of citizens in data collection and data analysis by crowdsourcing innovative ideas and data sharing through social media platforms.

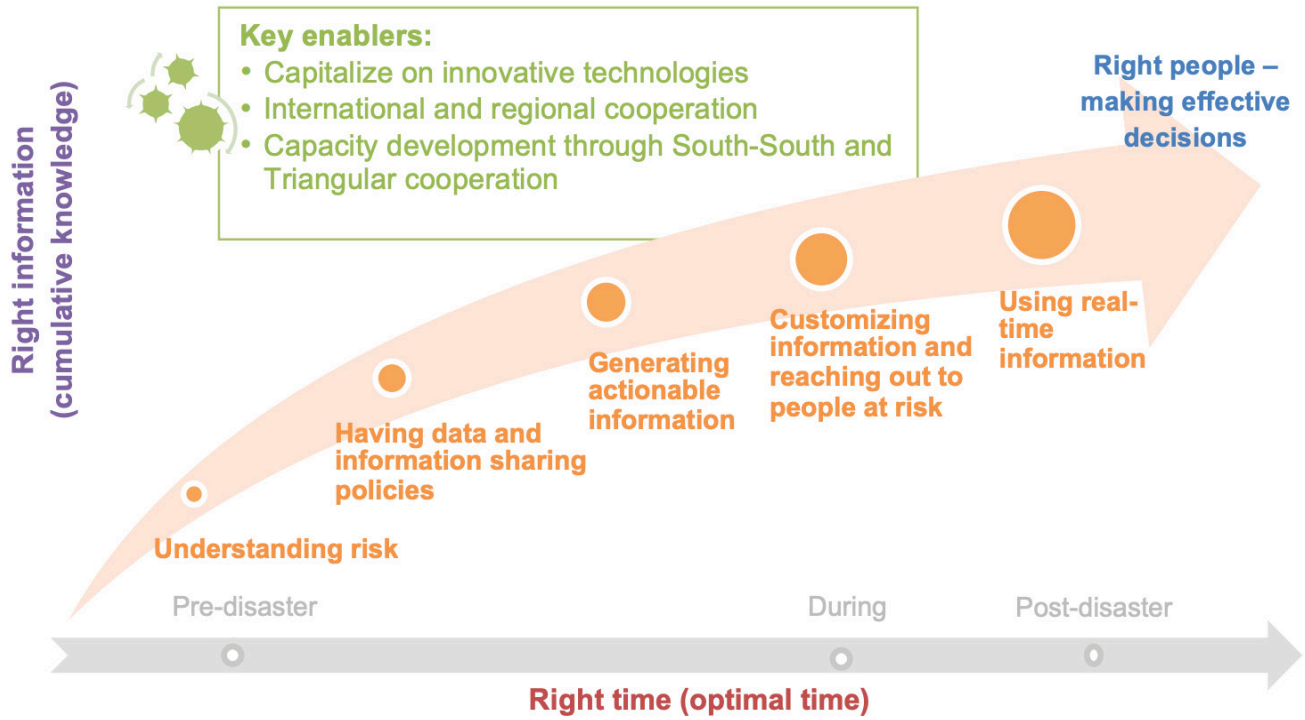
Data Security and E-Resilience of ICT infrastructure

During public health emergencies like COVID-19, many activities and services, such as healthcare, education, and employment, must be converted online, with higher dependence of societies on ICTs and digital government systems. This has led to a big surge in visits to online service infrastructure and digital government portals. In this regard, governments are required to take swift actions to ensure robust and seamless online systems so that citizens, particularly the most vulnerable groups, can still access essential public services and social benefits

10 https://corona.go.jp/

11 https://www.kaust.edu.sa/en/news/hacking-the-sars-cov-2-genome

Figure 3
Illustration of e-resilience guiding principles



Source: ESCAP (2015). Asia-Pacific Disaster Report 2015, p.131.

such as unemployment benefits during the times of public health emergencies.

At the same time, an integrated digital government system can become more vulnerable to cyberattacks and security breaches during emergency situations. Therefore, it is even more critical to strengthen data security and e-resilience of ICT infrastructure than during normal times. In the United Arab Emirates, in response to the global pandemic, the Telecommunications and Regulatory Authority (TRA) has launched several initiatives to enhance the ICT sector in response to the COVID-19 pandemic.[12] In addition to the initiatives on providing online services to citizens, TRA's initiatives include those for providing free internet data for health applications, increasing the speed of fixed broadband to ensure smooth running of online work and distance learning, securing IT systems and infrastructure, and raising security awareness among the general public.[13]

E-resilience is ICT contributions to resilience particularly at the community level.[14] In the case of public health emergencies, e-resilience entails two main dimensions: i) ICTs and digital government for meeting the temporary surge in demands on ICT-based and digital services; and ii) the capacity to maintain critical services and rapid restoration of ICT infrastructure and services in case of system failure. The Asia Pacific Disaster Report 2015 identified five essential steps and guiding principles to enhance e-resilience: understanding risk; installing data and information-sharing policies; generating actionable information; customizing that information and reaching out to people at risk; and using real-time information (*figure 3*).

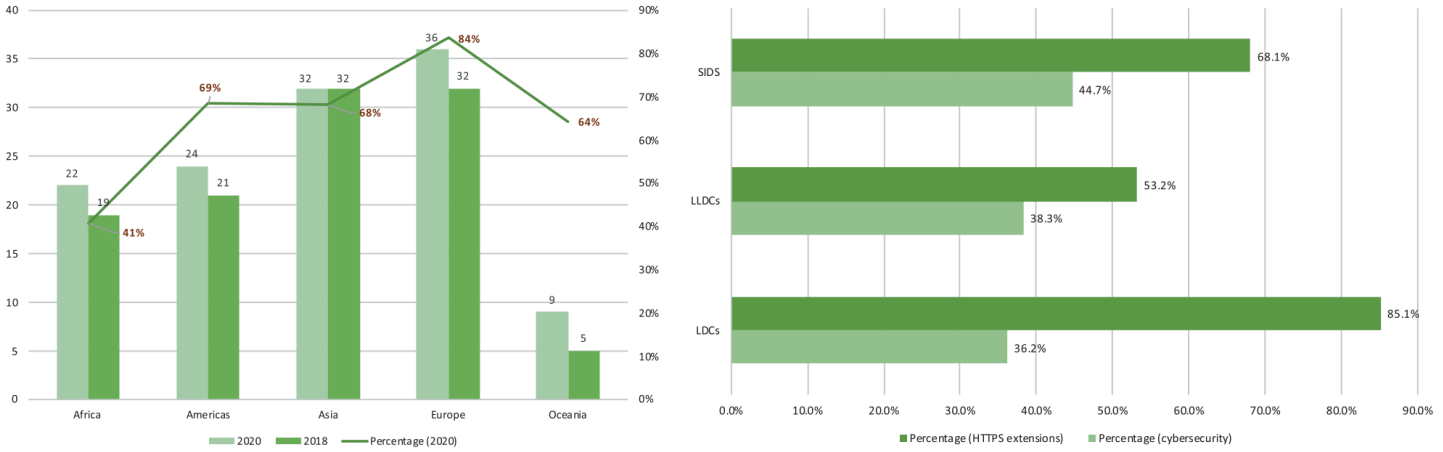
Many countries have suffered from some form of government data security breach. People are worried about their personal data being lost or stolen. Data security breaches not only impair the effective functions of institutions but also affect the safety and security of people. They also undermine public trust in government, while public trust is critical in times of public health emergencies.

To address public concerns over data security breaches, many governments have enacted cybersecurity laws for ensuring online data security and protection. The UN E-Government Survey 2020 shows that the number of countries with cybersecurity legislation available online has increased from 109 in 2018 to 123 in 2020, with the latter figure representing 64 per cent of Member States (*figure 4*).

12 https://www.tra.gov.ae/en/about-tra/tra-initiatives-in-response-to-covid-19.aspx
13 Ibid.
14 Heeks, R. & Ospina, A. (2018). Conceptualizing the Link Between Information Systems and Resilience: a developing country field study. Information Systems Journal.

Figure 4

Regions and country groupings with cybersecurity legislation available online and/or with HTTPS extensions in place



Key: SIDS, small island developing States; LLDCs, landlocked developing countries; LDCs, least developed countries.

Source: UN DESA (2020). UN E-Government Survey 2020, p. 161.

Inclusion of Vulnerable Groups in Data Collection and Data Disaggregation

It is often the case that vulnerable groups, such as people in poverty, people in rural and remote areas, women, children, indigenous peoples, and migrants and refugees, are under-represented or excluded in the process of data collection, while they are the groups that are most severely affected by the impact of COVID-19 due to their special vulnerabilities. Migrants, refugees and internally displaced persons (IDPs) warrant special attention during COVID-19 as they are often not included in the statistical reporting system. If people are not accounted for, they will not be eligible for government benefits. This is the case of homeless people or those working in the informal economy in many countries. One of the primary reasons for this type of exclusion is weak government capacities in civil registration and vital statistics, such as ID management systems.

In this regard, it is essential that government ensures that such vulnerable groups are included from the outset of the data collection process, especially when conducting a rapid assessment survey for selected local areas that are most hit. Typically, a rapid assessment is conducted immediately after the onset of a disaster in order to assess the disaster-affected areas and the needs of disaster victims.[15] Vulnerable groups must also be included in the processes of data dissemination and data classification. For example, in Cuba, a dashboard about the latest developments of the pandemic has been created.[16] It offers information such as the distribution of cases in the municipalities, disaggregated by sex, mode of infection, age, number of cases per day, recoveries and fatalities. Geographic information systems allow the performance of spatial analysis, network analysis, statistical analysis or information exchanges to show results on potential areas where to locate disease outbreaks, disease dispersal mechanisms, or identify optimal routes for the transfer of patients quickly and safely.

Data disaggregation is key during a pandemic.[17] Generating disaggregated data by gender, age, disability, etc. is imperative for accurate assessment of the situation and addressing inequalities by identifying the groups most affected and their special vulnerabilities and demands during the pandemic. Moreover, data disaggregation should be further refined to include more than one aspect, given the impact of intersectionality on socio-economic development. For example, the COVID-19 Sex-Disaggregated Data Tracker,[18] the world's largest database of sex-disaggregated data on COVID-19, tracks differences in COVID-19 infection, illness, and death among women and men. The National League of Cities in the U.S. also calls for data disaggregation in COVID-19 response, particularly by collecting and reporting COVID-19 cases by race.[19] While governments shall take measures to improve data disaggregation for the inclusion of

15 Maya Arii (2013) 'Rapid Assessment in Disasters', Japan Medical Association Journal, 56(1), pp. 19-24 https://www.med.or.jp/english/journal/pdf/2013_01/019_024.pdf

16 https://covid19cubadata.github.io/#cuba

17 Pan American Health Organization. https://iris.paho.org/handle/10665.2/52002

18 https://globalhealth5050.org/the-sex-gender-and-covid-19-project/

19 https://citiesspeak.org/2020/04/29/city-leaders-call-for-data-disaggregation-in-covid-19-response/

vulnerable groups in response to COVID-19, it is equally important to protect the privacy of vulnerable groups and avoid exposing them to potential discrimination and unwanted threats and harms.

Effective Risk Communication Strategy and Public Trust

The COVID-19, characterized by high rates of infection, significant morbidity, and rapid increases in cases, calls for an effective risk communication strategy. According to the World Health Organization (WHO), risk communication refers to the exchange of real-time information, advice and opinions between experts and people facing threats to their health, economic or social well-being. An effective risk communication strategy is instrumental for fighting rumours or misinformation, which may cause panic and generate anxiety and hoarding behavior, avoiding unfair treatment or discrimination against certain races and ethnic groups, and garnering public support to government response actions. Effective and timely data release by government is the key to risk communication to the public.

Most governments have created centralized data platforms and share up-to-date data and information through various channels, including government websites, social media, and mobile apps. A recent review of the 193 UN Member States' national portals showed that nearly 97.5 per cent (188 countries) had information about COVID-19 on their national portals.[20] Thorough and timely sharing of government data contributes to understanding of citizens about the COVID-19 pandemic and the performance and priorities of governments in combating the virus. This is critical for building trust in government. Open and transparent data sharing also contributes to strengthening accountability during times of emergency. For example, in Ukraine, the Ministry of Digital Transformation launched an initiative to inform people about the pandemic and to avoid the spread of COVID-19 through the Diya mobile app (a government mobile app for e-documents) with up-to-date information on the spread of COVID-19, quarantine conditions, government decisions and practical recommendations on how to prevent COVID-19 infection.[21]

Many governments have been taking measures in data sharing and risk communication to counter the spread of misinformation, which triggers fear among citizens and stigmatization of certain groups. UN Secretary-General António Guterres emphasized that "we are also seeing another epidemic, a dangerous epidemic of misinformation" and highlighted the importance of "trust in institutions—grounded in responsive, responsible, evidence-based governance and leadership" to combat the infodemic.[22] To respond to the proliferation of misinformation, a governance framework for COVID-19-related information sharing and communications on the online platforms, especially social networking platforms, is needed by setting ethical standards and legal frameworks.

Creating an Enabling Data Ecosystem for Effective Solutions to the Pandemic

The application of a number of technologies that enable innovative solutions to the COVID-19 pandemic, such as big data analytics, AI, cloud computing, robots, and blockchain, critically depends on quality data, open data, and effective data partnerships. As the role of these technologies in developing effective, timely, and inclusive COVID-19 responses is becoming increasingly important, so are the needs for accurate, quality, and secure data from various sources and use of data by many different actors. Such a high dependence on data, especially with an exponential increase in the amount of data being consumed and generated during a short period of the emergency, calls for creating an enabling data ecosystem through data governance.

Data ecosystem is one of the key pillars of data governance, reflecting the relationship between data processes and public engagement (*figure 5*). Although there is no agreed definition of data ecosystem, it generally refers to an open and dynamic system in which a number of stakeholders interact with each other to exchange, produce and use data. Such ecosystems provide an environment for creating, managing and sustaining data sharing initiatives.[23] A data ecosystem that enables effective data use during public health emergencies such as COVID-19 requires collaborative, cross-agency, and public-private partnerships, open data access and information sharing.

Common challenges in creating data ecosystem include lack of leadership or coordination mechanism,

20 UN DESA (2020). Compendium of Digital Government Initiatives in response to the COVID-19 Pandemic, p.1

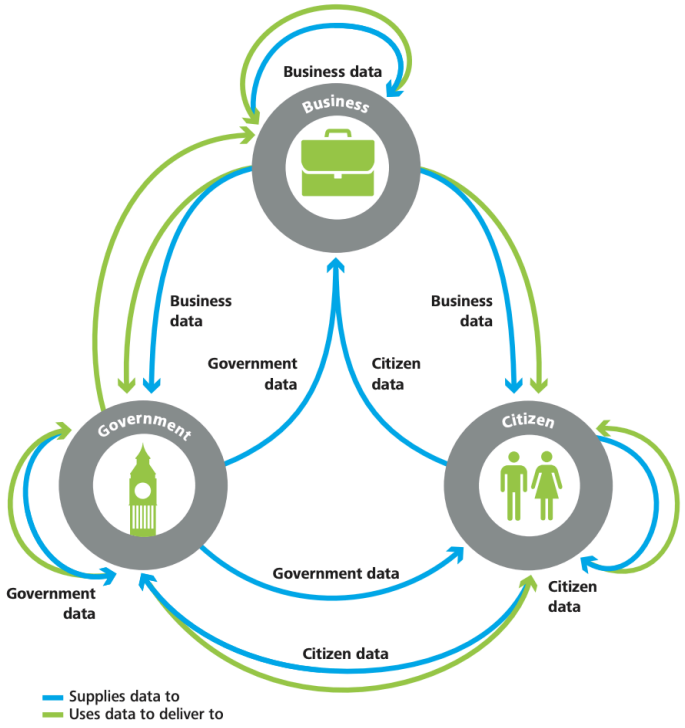
21 https://thedigital.gov.ua/news/mintsifri-zapuskae-push-povidomlennya-u-zastosunku-diya

22 UN Secretary-General's video message on COVID-19 and Misinformation, 14 April 2020 available at: https://www.un.org/sg/en/content/sg/statement/2020-04-14/secretary-generals-video-message-covid-19-and-misinformation

23 Oliveira, Marcelo and Lóscio, Bernadette. (2018). 'What is a data ecosystem?', Proceedings of the 19th Annual International Conference on Digital Government Research: Governance in the Data Age, pp. 1-9. 10.1145/3209281.3209335. https://www.researchgate.net/publication/325493490_What_is_a_data_ecosystem

Figure 5

The open data ecosystem



Source: Deloitte LLP (2012), Open Data: Driving Growth, Ingenuity and Innovation, p.9.

institutional barriers, lack of trust, poor data quality, lack of interoperability, and lack of organizational policies restricting data sharing. To overcome these challenges, national statistical offices (NSOs) shall play a greater coordination role in ensuring the timely and more open flow of information between different organizations—both official and nonofficial—within the statistical system at the national level, which in most cases requires legal agreement and governance measures.

Sixty-two per cent of countries have launched new data collection efforts to monitor the impact of COVID-19, and over fifty per cent have set up platforms to serve public data needs during the pandemic.[24] Government institutions, especially national statistical offices, are working in new ways to process, analyze, and communicate data. They are also increasing collaboration with non-governmental partners from the private sector and academia, which requires stepped up efforts to protect privacy and ensure data quality.

24 Statistics Division of UN Department of Economic and Social Affairs, Trends in National Data Ecosystems in Times of COVID-19. https://unstats.un.org/unsd/undataforum/blog/Trends-in-national-data-ecosystems-in-times-of-COVID-19/

RECOMMENDATIONS FOR INSTITUTING DATA GOVERNANCE TO ADDRESS THE COVID-19 PANDEMIC

- Governments should make every effort, through legal or technical means, to ensure the privacy and security of personal data, while instituting rigorous protocols for data partnerships between government institutions and businesses.
- It is important to adopt a holistic and whole-of-government approach to data governance with the engagement of all stakeholders and partners across sectors. Building data partnerships with all stakeholders can help leverage digital solutions driven by the private sector, promote publication of data produced by civil society organizations on open government data portals or open government data on non-government data portals, and support data sharing among all stakeholders.
- Governments at all levels need to build up their capacities to overcome data silos and skill gaps to address diverse dimensions of data governance. These range from ensuring the consistency of data collection to enhancing government accountability in sharing data and strengthening data quality and data security for a timely and proper response.
- It is imperative to enhance the capacity for the collection of disaggregated data to ensure the inclusion of vulnerable groups that are hit most during the pandemic due to their special vulnerabilities, while at the same time ensuring privacy protection for the members of these vulnerable groups.
- Data quality assurance frameworks[25] for timely, quality, and integrated data as well as the policy for data partnerships and interoperability should be developed and applied for data collection, processing, and dissemination.

25 United Nations National Quality Assurance Frameworks Manual for Official Statistics. https://unstats.un.org/unsd/methodology/dataquality/

REFERENCES

UN DESA (2020). COVID-19: Embracing Digital Government during the Pandemic and Beyond. Policy Brief No. 61. New York: United Nations.

UN DESA (2020). UN E-Government Survey 2020. New York: United Nations.

UN DESA (2020). Compendium of Digital Government Initiatives in response to the COVID-19 Pandemic. New York: United Nations.

UN DESA (2018). UN E-Government Survey 2018. New York: United Nations.