30 June 2020

Asia Internet Coalition (AIC) response to the Consultation on World Development Report 2021: Data for Better Lives

To
The World Bank
1818 H Street, NW Washington, DC 20433 USA
WDR2021@worldbank.org

On behalf of the Asia Internet Coalition ("AIC") and its members, we would like to submit our comments and recommendations **to the Consultation on World Development Report 2021: Data for Better Lives ("Consultation")**. AIC is an industry association comprised of leading internet and technology companies in the Asia Pacific region with a mission to promote the understanding and resolution of Internet and ICT policy issues in the Asia region. Our members are Google, Facebook, Amazon, Apple, SAP, LinkedIn, Booking.com, Airbnb, Cloudflare, Expedia Group, Grab, LINE, Rakuten, Twitter and Yahoo (Verizon Media). We have worked closely with the governments around the region in relation to the development of ICT policies and in doing so, we have witnessed first-hand the potential for adoption of digital technology and innovation.

With changing data landscape that will benefit people in low- and middle-income countries, the AIC commends the World Bank on leading with this Consultation. We also echo that integrating public and private data offers tremendous potential to improve the lives of poor people, but also opens backdoors that can harm people, businesses and societies. Against this backdrop, AIC would like to effectively contribute to improving the development outcomes through better public policies, program design and service delivery.

As such, please find appended to this letter detailed comments and recommendations, which we would like to respectfully request the World Bank to consider, which could be a useful feedback for future consultations and report development. The response to consultation is divided into two parts. While, Part 1 details out the Specific Comments in Response to the Concept Note and Part 2 provides answers to the consultation questions.

Should you have any questions or need clarification on any of the recommendations, please do not hesitate to contact us directly at Secretariat@aicasia.org or +65 8739 1490. Importantly, we would also be happy to offer our inputs and insights on industry best practices directly through virtual meetings and discussions to help shape the World Development Report 2021.

Thank you for your time and consideration and we look forward to hearing from you.

Sincerely,

Jeff Paine
Managing Director
Asia Internet Coalition (AIC)

Introduction

Asia Internet Coalition (AIC) welcomes the focus of the World Bank's *World Development Report* (WDR 2021) on the value of data for global development and sincerely appreciate the opportunity to provide comments to the Concept Note. The World Bank's effort to create a conceptual framework to guide countries and businesses towards more efficient and widespread practices of using data to foster socio-economic welfare, is timely and needed. Based on many years of delivering secure cloud infrastructure to tens of thousands of customers worldwide, including governments, commercial enterprises, not-for-profits, and universities, we have observed cloud services support the development of low and middle-income countries. Cloud enables organizations of *all* sizes to quickly scale up with minimum investment, bypassing many of the costs associated with traditional IT infrastructure. They also provide affordable access to on-demand, high-performance and secure IT services. We therefore share the Concept Note's assessment that cloud is an essential component of the infrastructure needed to maximize the benefits accrued from the analysis of data, and that ensuring technical capabilities to use such data is an immediate and critical global challenge to communities in less developed and emerging countries.

This submission sets out our feedback to the discussion paper. The first half sets out our specific feedback and recommendations in response to the Concept Note, specifically in relation to its characterization of the value of data and data and cloud infrastructure. The second half addresses the specific questions posed by the World Bank in the consultation document. In particular, we also suggest additional areas for the World Bank to consider in their report including environmental and energy considerations and the need for Cloud First Policies. Throughout, we seek to provide perspectives that point to an environment that best supports beneficial use and re-use of data for development.

Part 1: Specific Comments in Response to the Concept Note

1. **Quality of Data should be a Fundamental Framing Device for WDR 2021.** While the Concept Note rightly articulates the value and potential of data in helping to drive developmental outcomes, it makes a fundamental underlying assumption that more data equals to better insights and therefore better outcomes for developmental objectives. However, this assumption fails to take into account that the collection, use and repurposing of data does not occur in a vacuum, and depending on why and how data was collected in the first instance, the quality of that data could severely affect the ability to draw better insights and better development outcomes. This is to some extent addressed in the Concept Note on page 23 under "Risks of Challenges of using private data for development purposes". However, quality of data issues – i.e. whether the data is fit for a new purpose and whether data contains inherent unintended bias – is equally applicable to both public and private data. **We therefore recommend that the quality of data should be a fundamental framing device to how the WDR understands the value of data and accordingly how to extract that value**. In this regard, assertions such as those on page 5 and 6, that the accumulation of data leads

to concentration of power, or that the ability to repurpose private data for public policy making would lead to positive outcomes for development, should be appropriately reviewed through the lens that all this is contingent upon the fact that the data is of good quality and fit-for-purpose. Other considerations that can influence the value of data, include questions relating to access and collection, substitutability and replicability, quality and usability, of the data itself. Furthermore, misuses of data are not limited to scenarios where governments are seeking to serve a political agenda, where there are cybersecurity or data protection breaches. A significant risk that is not sufficiently articulated throughout the document is the re-purposing of poor-quality data or data that is unfit-for-purpose, leading to policy decisions which could have significant negative implications for the public.

2. **Recognizing Costs of Collecting, Processing and Using Data, the Value of Proprietary Data and Encouraging Voluntary Sharing.** The WDR's conceptual framework for the value must recognize that the forced sharing of data between the commercial sector and public sector must be balanced against the chilling effect to innovation and investment in the digital economy. The Concept Note characterizes all data as non-rivalrous, but excludable – and asserts that data can therefore be repurposed by many firms and governments at the same time at marginal incremental cost. This assertion fails to take into account the costs that are involved in collecting, collating and processing data, to ensure that data can then be analyzed for an intended purpose. In addition, it also fails to account for the importance of proprietary data in valuations and business models, and therefore any policy that forces the provisioning of private sector data will have serious adverse economic implications.

   The WDR should therefore encourage existing international practices of successful data sharing models. There are many good models that exist for voluntary data sharing. For example, voluntary frameworks for the management of community data such as the Maori Data Sovereignty Network in New Zealand and Canada's OCAP® (Ownership, Control, Access and Possession) Standards for Data Sovereignty of Indigenous People, enables sharing of data within a community for defined purposes. In contrast, **the WDR should eschew the model of forced sharing of data between the private and public sector as this is a wholly untested area, with potential far-reaching and unintended negative consequences, including to competition, employment, innovation and economic growth**. Forcing the sharing of data would devalue data sets used by businesses to improve and provide their services, thereby dampening data innovation and creating a hostile investment environment. More broadly, it could also dis-incentivize both domestic and foreign organizations from operating in a given country, as they will want to preserve the confidentiality of their proprietary data.

3. **Differentiating Between the Different Types of Data Infrastructure and Mischaracterization of the Cloud.**

The Concept Note also covers a breadth of digital infrastructure and lumps together telecommunications, data centers, software and hardware as "data infrastructure," which does not accurately reflect the complexities of each component. As a result of this broad-sweeping definition, some observations made throughout the Concept Note may not necessarily apply to all types of infrastructure as defined. Remote hosting and processing are core features of cloud and are necessary to enable the transformative capabilities that truly maximize the opportunities and value of data. The Concept Note however appears to mistakenly conflate international bandwidth costs with the costs of cloud services in its analysis. These are distinct issues, as the location of data centers does not curtail the ability of people from different income groups and geographies to access data infrastructure, which is largely determined by other factors, such as Internet access costs and government regulations. In other words, cloud services are actually a crucial enabler as it provides developing countries the ability to access state of the art, secure, on-demand IT resources via the internet, even if there is no physical data center in the country. Instead of proposing to "balance" the costs of international bandwidth against the lower costs for cloud services in international markets, the report should focus on solving the fundamental problem – the lack of affordable Internet access.

Furthermore, the Concept Note observes that data centers tend to be located in developed countries and that this contributes to leaving access to data infrastructure skewed towards higher income groups.[1] This type of characterization is counter-productive and risks pitting developed markets against developing countries, or foreign companies against local companies. The slant towards developing national-level data infrastructure also fails to consider issues such as security features that can only feasibly be achieved with scale, breadth and/or depth of services available to power innovation – in this case, international Cloud Service Providers. It is therefore important that customers in these countries have the ability to choose between domestic and international service providers depending on their needs and use cases.

4. **Appropriately Characterizing the need for Data Protection, Security and Trust.** While acknowledging that allowing free flows of data carries important socio-economic benefits, the Concept Note appears to suggest that this can come as a trade-off as "the very protections around data which create incentives for data sharing and therefore increase the value going into the data lifecycle."[2] This type of concern is further developed when the Concept Note alludes to "non-financial considerations" associated with the "protection and control of access to sensitive personal data", which seems to claim that cloud services fail to offer the necessary protection for sensitive personal data. We fully concur in recognizing the importance of a rights-based legal approach to

[1] WDR 2021 Concept Note, Page 28.

[2] WDR 2021 Concept Note, Page 31.

data protection, however, we think that the Concept Note has mischaracterized the trade-off. Specifically, where it fails to cite the crucial statement from its source (OECD 2019's *Enhancing Access to and Sharing of Data – Reconciling Risks and Benefits for Data Reuse across Societies*), that the "Group of Seven (G7) ICT Ministers explicitly stated that except for cases with legitimate public policy objectives, […] to oppose data localization requirements that are likely to hinder the free flow of information." Furthermore, this type of characterization suggests a dichotomy between cloud-based architectures and users' rights which, in our view, does not have a real justification. Cloud users retain full control and ownership over their data, including for what it pertains to identity and access controls. They can also determine where they would like to store their data geographically.

In addition, the Concept Note touches repeatedly on the importance of creating trust around the collection and use of data, citing the need for data users and collectors to be accountable to "data providers". From the outset, it is unclear what "data providers" means in this document – whether it is referring to an individual providing their data, or an organization that was the first collator of the data. The Concept Note also mischaracterizes the role that algorithm-use often plays in data analysis and analytics. Ensuring that algorithms which are used to process data are fair and unbiased is critically important to establishing and maintaining end-users' trust and to ensure that everyone has the opportunity to realize its benefits. It is, however, important to highlight that bias enters these systems because of bias that is in the data, and because the data values, inputs and algorithms are created by humans – who are also biased. In this regard, an important component in the trust equation is data accountability. **We recommend that the WDR highlights the need to build initiatives that promote responsible data management practices and compliance of data with the FAIR principles (findable, accessible, interoperable and reusable)[3] to build trust and ensure re-usability of the data**. This could include understanding the lineage of data, keeping a data provenance record and ensuring that bias is minimized the source data.

[3] These principles are promoted by LIBER Europe and provide guidance for scientific data management and stewardship and are relevant to all stakeholders in the current digital ecosystem. They directly address data producers and data publishers to promote maximum use of research data.

Part 2: Responses to the Consultation Questions

Responses to Question #1: Opportunities and Challenges to Improve Lives through Data.

The benefits of using and re-using data are truly unlocked when information is allowed to flow freely, including across national borders. We share the World Bank's conclusion that the free flow of data has a net-positive impact on the global economy. This assessment is supported by numerous studies that estimate significant costs connected to establishing barriers to data flows,[4] which are more arduous to bear for low and middle-income countries. A report by the McKinsey Global Institute estimates that cross-border data flows can contribute up to USD 11 trillion to the global economy by 2025.[5] The need to address the economic repercussions of the COVID-19 pandemic calls for global responsibility not to undermine this enabler of growth, and to guide less developed countries through a process of digital transformation. Essential verticals like finance and healthcare have quickly turned to cloud to build resilience into their operations. For some businesses, the shift to cloud goes beyond cost-saving, it becomes an existential need as a majority of the labor force works from home. The following are specific examples of how the Cloud has helped achieve this.

1.1. **Harnessing Data to Face Global Challenges**. The value of cross-border access to data goes beyond pure economic considerations, as the world's collective response to the COVID-19 pandemic increasingly shows. Cooperation among countries through data sharing is a vital instrument for detecting infections, mitigating their spread, and to identifying and making available treatments and vaccines. Unencumbered data sharing is supporting efforts to address the current crisis and promises to make countries better prepared to face future challenges. It is facilitating the biopharmaceutical industry, government and research institutions to share real-time information from clinical trials and from the screening of global libraries of medicines to identify potential treatments.[6] Beyond the current pandemic, data also enables disaster risk mitigation and remediation efforts through real-time mapping and use of AI and Machine Learning to more effectively identify and assess potential risks, as well as deploy resources for humanitarian relief efforts. Technology has the power to provide "first point of access" to populations in remote areas who may not have access to locally available resources, providing what were previously inaccessible public goods through data-based connectivity.

1.2. **Data for the Public Sector**. The Concept Note also seeks to develop evidence-based recommendations on how data collected for public intent can be most efficiently used to design, execute and evaluate public programs and policy. Public sector cloud integration has tremendous

[4] The European Centre for International Political Economy (ECIPE) studied the economic impact of data localization requirements in seven markets and found that unilateral restrictions on cross-border data flow and access to foreign markets negatively impacts economic growth and recovery because it limits access to competitive pricing, job growth in many services and goods sectors, and investment opportunities. The study noted that data residency requirements not only impact data flow, but also a broader set of commercial expansion opportunities that rely on cross-border data flows.

[5] McKinsey Global Institute (2020). Digital globalization: The new era of global flows.

[6] Pharmaceutical Research and Manufacturers of America, Biotechnology Innovation Organization, "Our Commitment to Beat Coronavirus."

potential to improve availability, quality, timeliness and policy relevance of public data, and ultimately enable better development outcomes, for example, through increasing transparency, oversight and accountability. On top of benefiting the operation and delivery of government services, cloud hosting significantly improves accessibility of public data, which leads to desirable outcomes in terms of empowering civil society and informing local business decisions, leading to democratic and socio-economic growth.

1.3. Data-Enabled Services and Applications to Spur Business Growth. On top of unlocking development opportunities by enabling access to public data, Cloud is a direct enabler of local business growth. The Concept Note correctly acknowledges that data-driven platforms have the capability to allow small and medium enterprises (SMEs) and underprivileged communities to access national, regional and global markets and value chains, spurring product and process innovation and ultimately leading to greater economic welfare.

Question #2: Balancing Data's Potential and Risk of Misuse.

The Concept Note does a great job of identifying critical focus areas to facilitate economic development through data driven innovation. At the same time, we believe that it presents views that require some level of clarification to ensure that policy-makers and businesses can access precise, evidence-based and actionable recommendations.

2.1. Data Protection and Security. As it relates to regulation, we support the ongoing initiatives of governments to make their data protection regimes more interoperable and do so in ways that protect user privacy whilst balancing an organization's legitimate need to process personal data. An important objective in data protection laws is to encourage better internal data governance and improve data security-hygiene practices in organizations, and international privacy standards that have proven efficacy should be widely recognized as demonstrating and adequate level of protection. The Concept Note alludes to the work undertaken through the APEC Cross-Border Privacy Rules system; and should also recognize other global initiatives such as the OECD's Guidelines on the Protection of Privacy and Transborder Data Flows, as well as regional frameworks, such as ASEAN's Framework on Personal Data Protection. In addition, governments should recognize that data protection laws should be distinct from other laws which on cybersecurity, cybercrime and law-enforcement to data. Within dedicated cybersecurity frameworks, provisions should also be aligned with proven global industry-backed methods of risk management as well, such as the ISO/IEC 27000 family of information security management system standards and the NIST Framework for Improving Critical Infrastructure Cybersecurity.

It is our view that the concern around control and access to personal data is largely due to a widespread, yet unjustified "perception". Having full ownership of the "stack," all the way from the building floor and walls to the software on the servers, made people feel comfortable that their data was as secure as possible. This rationale still exists for many governments. Nonetheless, this is a false belief. Regardless of the physical location, if IT systems are in any way connected to the Internet (or other multi-party networks), even indirectly, they are at considerable risk, as breaches do not require physical access to a server but instead

exploit lack of effectively implemented logical security controls. Cloud offers the best mechanisms to protect, detect, respond, and recover personal data, leveraging modernization and automation to always guarantee the highest standards of security. Research by Gartner and IDC estimated that public cloud infrastructure as a service (IaaS) workload will experience at least a 60% reduction in security incidents compared to those in traditional data centers.[7] In contrast, regulations that look to curtail access to cloud services in the name of security risk bearing the dual negative effect of precluding access to transformative digital tools while, in practice, decreasing security. The solution to this problem of perception should rather focus on improving user and government understanding of cloud, establishing direct discussions between stakeholders and service providers.

2.2. Data Classification Frameworks. Data classification is a foundational step in cybersecurity risk management. Data classification has been used for decades to help organizations make determinations for safeguarding sensitive or critical data with appropriate levels of protection. It involves identifying the types of data that are being processed and stored in an information system owned or operated by an organization. It also involves determining the sensitivity of the data and the likely impact should the data face compromise, loss, or misuse. To ensure effective risk management, organizations should aim to classify data by working backwards from the contextual use of the data and creating a categorization scheme that takes into account whether a given use-case results in significant impact to an organization's operations (i.e. if data is confidential, needs to have integrity, and/or be available). We recommend that public sector entities to adopt and encourage uptake of data classification models, starting with a three-tiered approach covering "Unclassified", "Official", and "Secret and above", which has shown to sufficiently meet both public and commercial customer needs and requirements.

Question #4: Policy Reforms to Harness the Value of Data in Developing Countries.

Based on our experience of partnering with governments and businesses in less developed countries, this section of our response showcases policy alternatives that have proven capacity to harness the value of data to support socio-economic development, while guaranteeing against the risks of its misuse.

4.1 Cloud First Policy. A cloud first policy directs government entities to use public cloud services as the default, and in preference to traditional IT systems. It aims to accelerate cloud adoption within government, helping public administrations make more efficient use of their data. Cloud first policies generally rely on public cloud models, which leverage the benefits of scale to realize significant cost savings for governments and better services to citizens. They improve the capacity for governments to more easily share data across agencies and with the public, while also ensuring the highest levels of security to public sector data. A positive example of this model is that of the Philippines, which implemented a cloud first policy in 2017 and were quickly able to achieve improved inter-agency collaboration and better citizen services. Public sector cloud integration allowed for flexibility, security, and cost-efficiency by leveraging global systems of solutions, innovations and services, as well as updated cybersecurity measures.

Furthermore, cloud native policies should be encouraged to focus on utilizing "born in the cloud" solutions, rather than simply directing government agencies to lift and shift their current IT operations to the cloud.

[7] Gartner (2019). Is the Cloud Secure.

In a typical cloud journey, a public sector entity starts by using cloud computing for basic infrastructure needs, such as compute and storage, and then makes a step towards cloud native by picking "born in the cloud" solutions. In the final ideal end state of a government with a cloud native policy, government and public sector entities will be opting for standalone "born in the cloud" solutions that maintain the similar benefits to those realized under cloud first.

4.2 Economic Policy Challenges. Throughout its length, the Concept Note presents concerns that the concentration of market power in global data-driven businesses may preclude entry to the data economy by small firms. Our members experience – and that of their SME partners – tells that strategic synergies between global players and local businesses, where the benefits of data-driven innovation are shared and small firms can scale, are a reality. Instead of imposing limitations on the structure and operations of large data-driven firms, we encourage countries to develop voluntary best practices that align to regional and international standards and codes of conduct. The chapter on competition should be careful not to oversimplify considerations for how possession of data leads to a competitive advantage. A wide range of criteria can influence how data can impact competition, including questions relating to access and collection, substitutability and replicability, quality and usability. Prohibitive rules around large companies may threaten innovation and contractual freedom, which in turn may adversely affect market development and harm consumers, particularly in less advantaged communities.

4.3 Open Data. Policy frameworks should take into account existing global best practices of successful data sharing models for the voluntary sharing data. There are many exemplar models that exist for voluntary data sharing. These datasets can be leveraged for a wide variety of solutions, such as public sharing, research, testing, large-scale data analysis, and archiving; and there are a wide variety of scenarios and purposes in which personal and non-personal data could be shared calls for flexible and voluntary frameworks. We would therefore encourage the World Bank to leverage these global examples and consider how best to incentivize voluntary data sharing.

Question #5: Other Issues.

5.1 Energy Efficiency of Data Centers. The environmental impact resulting from the processing of data, should also be developed as a consideration in this WDR. The International Energy Agency (IEA) has been calling for governments to closely consider resiliency and sustainability when making choices that will guide their COVID-19 recovery efforts and improve development outcomes. Cloud is a tool that governments can leverage to build in resilience while minimizing the impact on the environment, and when selecting a solution, organizations should consider resource utilization and energy efficiency, in addition to power mix. A typical large-scale cloud provider achieves approximately 65% server utilization rates versus 15% on-premises, which means when organizations move to the cloud, they typically provide less than a quarter of the servers than they would on-premises.[8] In addition, a typical on-premises data center is 29% less efficient in its use of power compared to a typical large-scale cloud provider that uses world-class facility designs, cooling systems, and workload-optimized equipment.[9] This has significant environmental

[8] NRDC 2014 "Data Center Efficiency Assessment" report.

[9] Power Usage Effectiveness (PUE) of on-premises data centers from 2014 Uptime Institute study and PUE of cloud data centers from Google and Facebook public disclosures plus AWS internal data, all of which show PUEs under 1.2.

benefits, especially considering that global data center energy demand can remain nearly flat through 2022 despite a 60% increase in service demand, if current trends in efficiency of hardware and data center infrastructure can be maintained.[10] Moreover, for governments and businesses, ensuring redundancy in the event of operational failure due to a disaster or other circumstances is integral for stability. Having clustered operations in only one country exposes the organization to a level of risk that can far outweigh data access concerns.

5.2 Capacity Building and Skills. The Concept Note acknowledges that realizing the opportunities of data goes beyond improving the physical data infrastructure. Closing the global skills gap has the potential to add USD 11.5 trillion to global GDP by 2028.[11] Public and private sectors have a joint responsibility to improve the human capital in terms of data literacy and skills, and public. For example, in Indonesia leveraging the technology expertise of cloud service providers on cloud skills enhancement initiatives to empower hundreds of thousands of Indonesians from all backgrounds with cloud services. Such programs are run in collaboration with the government agencies, various educational institutions, and other collaborators.

Conclusion

We support WDR 2021 as a guide to how governments, civil society, and businesses can maximise the benefits of data. While information sharing has always been a driver development, it will increasingly prove a vital ingredient in overcoming the unique challenges facing the world today. Crafting the right environment, one that considers and harmonizes the needs of various stakeholders, requires deliberate review. We are optimistic that WDR 2021 will take into account international best practices, consider proven policies such as data classification and cloud first policies, and provide objective views of various technologies that leverage data. In this regard, debates over the perception of security, other non-financial considerations, and competition are important and should be dealt with recognizing their complexity.

We believe that WDR 2021 can help countries optimize decision making to ensure that their people can benefit most from data, and we are grateful for the opportunity to contribute to this critical instrument for development.

[10] EA analysis based on Masanet, E. et al. (2020). Recalibrating global data center energy-use estimates, *Science*, 367(6481), 984-986, https://doi.org/10.1126/science.aba3758.

[11] World Economic Forum, "Closing the Skills Gap Accelerators," https://www.weforum.org/projects/closing-the-skills-gap-accelerators