Tackling online misinformation while protecting freedom of expression

Dialogue 2 in the Series "Frontiers of Digital Economy"

Report of discussions of the Expert Round Table on "Tackling online misinformation while protecting freedom of expression"

11 Oct 2021, 0830 – 1030 UTC Via Zoom

Executive Summary

Misinformation, broadly defined, is a challenge that governments, civil society, academia and industry are seeking to address. The purpose of the Dialogue was the sharing of experiences with various approaches to the problem so that all stakeholders, including governments, would be better informed when developing solutions in their national contexts.

Actions to limit the dissemination of misinformation must always balance the protection of freedom of expression with the avoidance of the harm. Misinformation is not a creation of social media, having been common throughout history, especially since the rise of the penny press in the 19th Century. However, the lowering of the transaction costs of sharing information and the speed by which any kind of information may spread has increased the salience of the problem.

While many civil society groups and even some government agencies are focusing on fact checking and fostering abilities to discern factual from inaccurate information, some governments have considered more draconian measures. It is therefore opportune to draw from experiences with the various approaches from around the world, keeping in mind the goal of minimizing harm while protecting citizens' fundamental rights, particularly the freedom of expression.

Codes such as the Australian Code of Practice on Disinformation and Misinformation have potential as approaches that permit stakeholders to adapt and improve solutions as the digital ecosystem evolves. The Australian regulator acting through DIGI, which was responsible for the Code, took special precautions to protect the freedom of expression as an implied right of political communication. For example, 'obvious' political speech was exempted. However, this was a limited exemption: if false statements by a politician are then artificially amplified through an orchestrated disinformation campaign, action would be taken. The importance of transparent reporting was highlighted.

Facebook's Remove-Reduce-Inform strategy, whereby proportional responses are given to misinformation, depending on severity, was also discussed. However, judgements as to what is misinformation and the degree of severity may be questioned, requiring the cooperation of fact checkers and other key stakeholders. For example, in the context of the COVID-19 pandemic, collaboration with global health experts was crucial in identifying harmful, false information. As part of the "inform" component various digital literacy initiatives have been undertaken in multiple countries, in collaboration with civil society organizations. The challenges of reaching the right audiences at scale were discussed.

The role played by Facebook's algorithm with regard to reducing the visibility of misinformation was examined. Educating and empowering users on the options to customise and control their News Feed, and prioritize accounts was considered. If the algorithm is removed entirely as some have suggested, misinformation may increase. Some participants expressed the wish that more could be done to provide reasons for why certain posts were sanctioned.

Digital literacy was seen as an essential complement to the various other approaches. Here, scalability is of paramount importance. "Train-the-trainer" models that focus on teachers and locally adaptable, tested content were emphasized. The opportunity of including digital

2

literacy in school curricula was underlined along with the responsibility of mainstream media in this regard.

Misinformation is a multi-stakeholder problem wherein collaboration with industry partners and companies can be very productive, as exemplified by the UK's effort over the last 4 years. Each actor in the ecosystem must leverage their own strengths. Governments may not have the capacity to react as speedily as platforms in terms of taking down harmful content.

When tackling misinformation, multiple objectives must be balanced such as public policy, market incentives, technology imperatives and fairness, and even how state actors themselves manipulate mainstream information flows, using multiple proxies. The debate on sovereignty in relation to digital platforms also must be explored further.

Introduction

Online misinformation is disseminated on a variety of topics by different actors of different motivations, through various means. Using the dimensions of harm and falsity, a report commissioned by the European Commission differentiated between three types of problematic information:

- Misinformation is false information that is often shared unintentionally.
- Disinformation is false information that is shared with the deliberate intent to mislead as part of a manipulation campaign or information operation.
- Mal-information is genuine information that is shared to cause harm, including private or revealing information that is spread to harm a person or reputation[1].

However, it is not always possible to determine intention, making the distinction between misinformation and disinformation somewhat problematic in a legal context. Disinformation also evokes notions of state-led dark operations, which is perhaps an artifact of its origins in the Russian KGB's dezinformatsiya operations. Therefore, the term misinformation as inclusive of disinformation is favored in this report.

Several governments and regional bodies have conducted extensive public hearings and/or commissioned reports, and, in some cases, enacted legislation dealing with all or some of the above-described categories of problematic information. Throughout history, information has been abused in various ways. With the rise of mass media in the 19th Century, in the form of the penny press that gave primacy to the assembling of large audiences for sale to advertisers, the incentives to attract attention by whatever means grew stronger[2].

As use of the internet has increased, it is imperative that all concerned tackle the emerging challenges based on dispassionate assessment of evidence. These challenges are not new or unfamiliar to society. The novelty is about how they are manifested on the internet and how best to respond to these manifestations. Part of the concern is the normal reaction to anything new. But a part is about the new elements introduced by social media. Many governments and stakeholders have expressed concerns about false information and the potential of harm, and sought remedial action.

While many civil society groups and even some government agencies are focusing on fact checking and fostering abilities to discern factual information from falsehoods, some governments have been attracted by more draconian measures. Such discussions are ongoing in the BBNMAPS (Bangladesh, Bhutan, Nepal, Maldives, Afghanistan, Pakistan, Sri Lanka) countries. It is therefore opportune to draw from experiences with the various approaches from around the world, keeping in mind the goal of minimizing harm while protecting fundamental rights of citizens, particularly the freedom of expression. Codes such as the European Union's Code of Practice on Disinformation[3] and the Australian Code

---

[1] Wardle, C.; Derakhshan, H. (2017). *Information Disorder: Toward an interdisciplinary framework for research and policymaking*. European Union, p. 20. https://rm.coe.int/information-disorder-toward-an-interdisciplinary-framework-for-researc/168076277c

[2] Wu, T. (2016). *The Attention Merchants: From the daily newspaper to social media, how our time and attention is harvested and sold*. London: Atlantic, pp. 17-18.

[3] European Commission (2018). *EU Code of Practice on Disinformation*. https://digital-strategy.ec.europa.eu/en/policies/code-practice-disinformation

of Practice on Disinformation and Misinformation[4], may be considered as approaches that permit stakeholders to adapt and improve solutions as the digital ecosystem evolves.

It must be noted that the recent publicity around the US Senate's hearings (during the first week of October, just before the Dialogue) was not about misinformation and its variants per se, but on the alleged disregard of research about the effects of the algorithm that determines what appears on the news feed. The focus of this discussion is on remedial actions that may be taken to minimize the harms from content propagated by various parties using social-media platforms. While keeping the focus on the topic of tackling misinformation while protecting fundamental rights including freedom of expression, the harmful or not effects of the algorithm are also discussed.

To understand how these debates are playing out in and affecting the BBNMAPS, this virtual dialogue was organized by LIRNEasia with resource persons from government, the private sector, academia and civil society to engage in frank roundtable discussions with invited decision makers from the BBNMAPS. This event was held on 11 Oct 2021, 0830 - 1030 UTC via zoom.

 The panelists were;

- Hon. Dullas Alahapperuma, M.P., Minister of Mass Media, Sri Lanka

- Dr. Jennifer Duxbury, Director Policy, Regulatory Affairs, and Research, Digital Industry Group Inc., Australia

- Saad Hammadi, South Asia Campaigner at Amnesty International

- Alice Budisatrijo, Head of APAC Misinformation Policy, Facebook

- Qadarauddin Shishir, Fact Check Editor for AFP, Bangladesh

The panel discussion was moderated by Rohan Samarajiva, founding Chair of LIRNEasia and former Chair of Sri Lanka's ICT Agency.

The invited experts included government officials and regulators, researchers, journalists, fact checkers, digital rights advocates, and legal experts, and representatives from the private sector. See aggregate attendance data in Annex 1.

This report summarizes the key points discussed, and offers recommendations on tackling online misinformation which we hope will be of use to policymakers in the BBNMAPS, as they work towards developing their own frameworks and legislation.

---

[4] Digital Industry Group Inc. (DIGI). (2021). https://digi.org.au/wp-content/uploads/2021/02/Australian-Code-of-Practice-on-Disinformation-and-Misinformation-FINAL-PDF-Feb-22-2021.pdf

Themes of Discussion

Definitions and Frameworks

Frameworks and definitions provide the foundation for tackling online misinformation. The importance of understanding contextual factors, especially the local digital ecosystem, when developing frameworks and policies was highlighted in the discussion. The Australian Code of Practice on Disinformation and Misinformation (prepared by DIGI, the Digital Industry Group) which was implemented in February 2021 was developed specifically for the local Australian context.

Australia was described as having a strong democracy with respect to freedom of expression and an implied right of political communication. Therefore, DIGI was instructed by the Australian Communications and Media Authority (ACMA) and the Australian government to be careful about how that right would be protected within the Code. DIGI decided to exempt 'obvious' political speech, an exemption which only applied to misinformation. For instance, if false statements are made by a politician which are then artificially amplified by an orchestrated disinformation campaign, action would be taken.

The Australian Code of Practice on Disinformation and Misinformation is a self-regulatory instrument. The perspective of the Australian Communications and Media Authority was taken into consideration in this regard. The broad range of signatories the Code has attracted, including both large and small platforms, has been made possible by its flexibility in terms of allowing signatories to choose the measures to tackle dis- and misinformation that are in accordance with their different business models and technological capabilities.

The Code was said to present additional features which distinguishes it from the European Code. There are two mandatory commitments that all signatories must abide by:

1. Implementing safeguards against dis- and misinformation

2. Filing annual transparency reports detailing how they plan to meet their Code commitments along with information about their policies

In October 2021, DIGI launched an oversight mechanism and a complaints facility to ensure signatories comply with their Code commitments. Apart from building on the first draft of the Code, however, the focus of DIGI also included the perspective of the platforms themselves.

When it comes to definitions, DIGI has defined mis- and disinformation as "False and misleading digital content that is disseminated on platforms and may cause harm." In both cases, the threat of harm must be serious towards society and at the collective level, including, but not limited to, democratic processes, vulnerable groups or social categories such as health and the environment. DIGI does not believe that it is feasible to determine the intentions of users spreading fake information. However, they distinguish between dis- and misinformation by the way that it is distributed. For example, disinformation is distributed by malicious actors who use strategies such as trolls or "sock puppet accounts" to seek to artificially amplify messages online.

Actions

Many stakeholders are involved in tackling online misinformation. Different roles are performed by governments, regulatory bodies, social media companies, traditional media and civil society organizations.

It is important that companies involved in moderating content uphold their human-rights responsibilities by engaging in human-rights due diligence and by ensuring greater transparency over content moderation policies. Technology companies have a major role in limiting misinformation and surfacing reliable information. How the US presidential campaign was subject to relatively more stringent fact checking mechanisms by the media was cited as an example. The use of independent fact checkers such as AFP by Facebook was commended.

As an example of the role played by regulatory bodies, the impact of transparency reports as part of 'Australian Code of Practice on Disinformation and Misinformation' was highlighted which helped provide insights into the different strategies introduced by platforms other than simply taking down content, for example, developing hubs of credible content and official information. Flagging and deprioritizing content in user feeds along with media literacy efforts are also quite important.

Misinformation is an issue taken seriously by Facebook. The complexity of phenomenon means that it is not feasible to remove all misinformation from the platform. There is a spectrum of content which different people consider to be false news which results in a lack of consensus on its exact definition. Facebook does not see its role as the arbiter of what is true or false.

Facebook employs different approaches to tackle the range of problems. To address disinformation, Facebook investigates the actors and their behavior. When the platform identifies attempts to deliberately share false narratives, all the accounts involved are taken down.

However, misinformation is often spread by people who genuinely think that what they are sharing is true. Therefore, Facebook does not wish to disproportionately penalize such users. Hence, the platform utilizes a three-pillar strategy:

1. Remove – Misinformation which could contribute to physical harm such as rumors and hoaxes that could incite violence, or even harmful misinformation related to COVID-19 and vaccines. Facebook also prohibits "deepfakes" (videos created by Artificial Intelligence which mislead viewers).

2. Reduce – For misinformation which does not contribute to harm, once verified by fact checkers, the distribution of the content will be reduced so that fewer people are exposed to it. For accounts that frequently share false information, all their posts are distributed on a much lower scale and their accounts will be removed from recommendations on the feeds of other users.

3. Inform – This is a critical component of the strategy as it aims to build a strong, digitally resilient society in the long term. By providing people with more context, and through investments in digital literacy, the aim is to equip people with the critical

media skills they need to decide what is true or false themselves. The platform continues to collaborate with community organizations to offer digital literacy programs, such as 'We Think Digital,' and 'fightcovidmisinfo.com' which are aimed at assisting people to think critically before sharing information. Ads were also run-on Facebook and Instagram in 27 countries, reaching 400 million people, to educate them on identifying misinformation related to COVID-19.

This approach balances the tradeoffs between keeping people safe while providing them with a platform to express themselves. However, Facebook does not believe they alone should be making so many decisions about content and since 2019, the platform has been calling on governments to play a role by regulating issues such as harmful content and protecting the fair conduct of elections.

Effectively halting the spread of misinformation requires collaboration from all stakeholders and authorities have a role to play in providing reliable sources of trustworthy information, especially in light of the ongoing infodemic. Social media companies should take their responsibilities toward countering misinformation seriously.

Initiatives taken by platform providers were appreciated but the need for greater effort was noted. Regarding the role played by Facebook to limit amplifications, several measures were mentioned. Systems are in place to reduce the spread of disinformation on the platform and the company shares updates of its efforts via a [monthly report](#) in the company's Newsroom.

Separate to this and with regards to misinformation, Facebooks' collaborates with relevant experts in removing misinformation that violate policies related to causation of harm.

The role played by Facebook's algorithm with regard to reducing misinformation was a topic of discussion. Facebook already gives people a lot of control over how they design their News Feed and users can mark certain accounts as their favorites, so they get content from them first. If the algorithm is removed entirely as some have suggested, misinformation will increase. This is because the algorithm is made with principles that embody integrity and keeps evolving to detect fake or low-quality information (such as spam or clickbait). The algorithm was also mentioned to have a system for reducing clickbaits, looking at content which is sensational that is on the borderline of Facebook's misinformation policies. Thus, removing the algorithm may make the issue worse.

On the topic of transparency, it was noted that there is demand for actions to increase transparency over why accounts have been removed or why content is flagged, thereby enabling users to understand how the platform deals with misinformation.

Steps that can be taken by governments to control the flow of misinformation were mentioned. Firstly, the Government can improve the speed with which the government reacts to misinformation, highlighting how a slow response could affect a large proportion of the public who may have been exposed to false information. The example of Sri Lankan Government's Health Promotion Bureau which has become an effective platform with a significant reach to distribute credible information was mentioned.

In terms of improving digital literacy, Facebook supports the process by showing clear warning labels and a debunking article from the fact checker for information flagged to be

fake. A notification is also given in the instance that content shared with friends in the past is proven to be false. Another product is a content agnostic feature that is used by Facebook to tell people that pictures or videos are out of context.

It was emphasized that helping people think critically of news and media, could be very effective in terms of enabling them to differentiate between true and false content. In the discussion, the significance of improving media literacy to counter misinformation was noted. The situation in Afghanistan was taken as an example. It was admitted that in situations like Afghanistan, fact checking can be challenging due to the low digital and media literacy levels of the general population. Additionally, it was noted that there has been extensive fact checking of content related to Afghanistan especially during the past few months. Another aspect to note is that misinformation originating from Afghanistan does not only affect people within its own jurisdiction but in other regions as well.

It is important for citizens to be diligent when seeking such information and validating the authenticity, credibility and veracity of such content. People need to make sure that they subscribe to information that is from credible sources, and not share information that they have not personally verified. The freedom to seek, receive and impart information needs to be transparent and whenever barriers arise in such situations, then people are required to make intelligent judgements over the content they come across.

Facebook has initiated numerous media literacy programs, both online and offline. Tutorial programs like 'We Think Digital' trains people in the Asia Pacific, a region where large populations are coming online for the first time, hence such programs serve to educate people about thinking critically before sharing information on the internet.

The need to ensure that media literacy programs are scalable and address the problem of information asymmetry was discussed. As a solution, Facebook is pursuing the model of 'train the trainer,' having reached roughly 1 million people in the Asia-Pacific. For instance, by training teachers, the model becomes more scalable as this knowledge of media literacy will have a greater reach, especially to the next generation. In this regard, the platform also calls upon governments to put digital literacy in the school curriculum from elementary school upwards, making it similar to an essential skill. The platform is also trying to target other demographics like senior citizens by working with community organizations.

The pandemic has provided much insight on information dissemination. It has shown that greater access to information can have varying effects but the access to credible information is most important. However, a key aspect to consider is the time sensitive nature of information, especially for COVID-19 related information such as data regarding vaccine efficacy, etc.

Exposing actors involved in disseminating false information was mentioned as important. Individuals and organizations dedicated to spreading disinformation campaigns are a challenge. If they are exposed to the public, the general population would be more likely to be skeptical when encountering related content.

Challenges

It was noted that the freedom of expression is fundamental to human rights and is also a precondition of democracy. The importance of governments having a progressive point of view when confronting the problem of misinformation without curtailing the right to free expression was highlighted.

Developing regulatory interventions to tackle misinformation while ensuring an appropriate level of protection for freedom of speech requires careful thought being given to the scope of the regulation as well as the mechanisms. For example, in Australia, because the media acts as a watchdog for the government and is already subject to self-regulation, news and media content that is subject to a published editorial code is exempted. However, if media or political content is artificially advertised by an orchestrated disinformation campaign, then it is included.

Government measures should not be through means of criminalization and in any event must follow the relevant international standards. Research by Amnesty International showed that some countries have adopted the pretext of public safety to suppress information, curtail discontent or crack down on the media and activism. Blanket restrictions on the dissemination of information based on vague and ambiguous concepts such as fake news, rumor, false information, or spreading misinformation are incompatible with international human rights law and standards. Further criminalizing false news is also in clear violation of the right to freedom of expression and risks creating a chilling effect on the general population and media, leading to self-censorship in fear of reprisal.

It is very important that states take responsibility to enact laws that address violations of the freedom of expression. It is also important to make provisions regarding breaches of privacy by state agencies, rather than simply criminalizing behavior. A cautious view is needed, as laws and legal structures could be weaponized against dissent. It was specified how authorities in Bangladesh, for example, use fake information as a pretext to target opposition. If legal steps are taken, they should be fair and impartial.

It was mentioned that misinformation and conspiracy theories could become more entrenched due to influential persons such as politicians contributing to the growing lack of trust in institutions. It was further highlighted that in the absence of sufficient credible information, misinformation fills the gaps, leaving those distrustful of their governments even more disinclined to accept official information. The importance of authorities addressing the spread of information that amounts to advocacy of hatred and incites discrimination was mentioned. For instance, publicly calling out persons with COVID19 creates risks to their safety.

The critical role conventional media play in informing the public was noted. For example, although Bangladesh has high internet penetration, traditional media still plays an important role (TV and newspapers). Thus, the media should play a larger role of fact checking and debunking false information in combination with online efforts. On the other hand, the media's capacity to operate should not be restrained, nor should journalists be penalized for their legitimate activities.

Misinformation is a multi-stakeholder problem where all actors need to play a role. Closer cooperation would catalyze actions to reduce the effects of false information. Any approaches taken to tackle misinformation should be conducted in a collective and strategic manner which values collaboration with industry partners and companies within the technological sphere such as the Asian Internet Coalition. The nuanced approach adopted by the UK which over the last 4 years, engaged in close cooperation with social media companies was mentioned as a model. Each actor in the ecosystem has their own strengths and weaknesses. Governments may not have the capacity to react as quickly compared to a platform in terms of taking down harmful content. Thus, the state should work in close cooperation with industry while abiding by agreed upon principles.

When tackling misinformation, many objectives that need to be balanced such as public policy, market incentives, technology imperatives and fairness, and even how state actors themselves manipulate mainstream information flows, using multiple proxies. The debate surrounding sovereignty and its connection to digital platforms also must be explored further.

Recommendations

Collaborative action to address misinformation is important

Policies and laws should ensure that freedom of expression is protected

Development of frameworks and policies should consider contextual factors such as level of democracy, freedom of press and attitude towards media literacy and awareness by the general public.
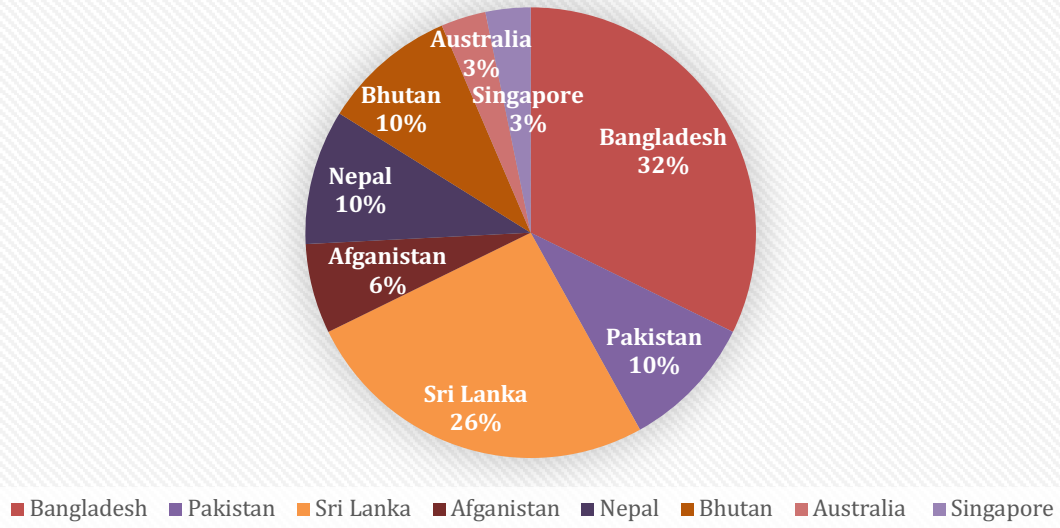
Human-rights due diligence and high transparency should be maintained when social media companies engage in content amplification, moderation and removal.

Annex: List of Participants

Panelists		
Hon. Dullas Alahapperuma, M.P.	Minister of Mass Media, Ministry of Mass Media - Sri Lanka	Sri Lanka
Dr. Jennifer Duxbury	Director of Policy, Regulatory Affairs & Research	Australia
Mr. Saad Hammadi	South Asia Campaigner, Amnesty International	Bangladesh
Alice Budisatrijo	Head of APAC Misinformation Policy, Facebook	Singapore
Qadaruddin Shishir	Fact Check Editor, AFP	Bangladesh
Participants		
Muhammad Solyh Ahmed	Fellow, Asia Internet Coalition	Bangladesh
Shahzeb Mahmood	Associate, Syed Ishtiaque Ahmed and Associates	Bangladesh
Faheem Hussain	Assistant Professor, Arizona State University	Bangladesh
Saimum Reza Talukder	Senior Lecturer, School of Law, BRAC University	Bangladesh
Ayesha Towhid	Bangladesh Institute of International and Strategic Studies (BIISS)	Bangladesh
Abu Saeed Khan	Senior Policy Fellow, LIRNEasia	Bangladesh
Prof. Dr. Shameem Reza	Mass Communication and Journalism Department, University of Dhaka	Bangladesh
Mr. Tarique M Barkatullah	Director, National Data Center	Bangladesh
Usama Khilji	Director - Bolo Bhi	Pakistan
Atif Aziz	Data Scientist - Islamabad Policy Research Institute	Pakistan
Haroon ur Rasheed Baloch	Senior Program Manager at Bytes for All (B4A)	Pakistan
Nalaka Guanwardana	Media thought leader and science writer - Independent	Sri Lanka
Ranga Kalansooriya	Regional Advisor Asia - IMS	Sri Lanka

Ashwini Natesan	Consultant / Lecturer, Tech. Media and Communications Law	Sri Lanka
Yudhanjaya Wijeratne	Watch Dog/Fact Checker	Sri Lanka
Mahoshadi Peiris	FactCheck.lk - Fact Checker	Sri Lanka
Stephanie Nicolle	Analyst - Verite Research	Sri Lanka
Dr. Davood Moradian	Director Afghanistan Institute for Strategic Studies	Afghanistan
Babu Ram Aryal	Lawyer	Nepal
Samiksha Koirala	PhD, consultant to Ministry on misinformation related matters	Nepal
Aadesh Khadka	Under Secretary, Ministry of Communication & IT	Nepal
Sangay Zangmo	Market and Competition Division, Bhutan InfoComm & Media Authority	Bhutan
Ugyen Dema	BICMA	Bhutan
Vindana Ariyawansa	Ministry of Mass Media - Sri Lanka	Sri Lanka
Dechen Chhoeden	Head, BtCIRT, Department of Information Technology	Bhutan
Tamim Asey	Advisor & Research Fellow, Afghan Institute of Strategic Studies	Afghanistan

Participant Countries

Bangladesh 32%
Pakistan 10%
Sri Lanka 26%
Afganistan 6%
Nepal 10%
Bhutan 10%
Australia 3%
Singapore 3%

Bangladesh · Pakistan · Sri Lanka · Afganistan · Nepal · Bhutan · Australia · Singapore

Participant Affiliations

Government 29%
Civil Society 39%
Private Sector 3%
Academic 29%

Government · Civil Society · Private Sector · Academic