A dialogue on transparency, open access and ethics in development research

3ie and Sehgal Foundation joint event, 4 December 2018, New Delhi, India

Sayak Khatua
International Initiative for Impact Evaluation (3ie)

Event report

April 2019

Note to readers

The International Initiative for Impact Evaluation (3ie), along with the Sehgal Foundation, hosted a one-day event on transparency, open access and ethics in development research on 4 December, 2018 in New Delhi, India. This report is a summary of the discussions that took place during the event.

This report has not been professionally copy-edited. Any errors or omissions are the sole responsibility of the authors.

Please direct all comments or queries to skhatua@3ieimpact.org.

Suggested citation: Khatua, S, 2019. *A dialogue on transparency, open access and ethics in development research*. New Delhi, India. International Initiative for Impact Evaluation.

Background

Research transparency covers the steps researchers take to put every aspect of their studies in the public domain. To promote transparency and openness, we need to advocate for improved accessibility; facilitate replication of processes and aid in theory-building. Transparent research practices increase the credibility of evidence generated. Openness enables users to access research materials such as data, study design, analysis plans, and use it according to acceptable and accountable ways that keeps a check on original researchers. This ensures that researchers and other users, especially in developing countries, have quick and affordable access to research, it improves knowledge sharing and democratises research.

Despite numerous efforts by a number of governments, international agreements and research bodies to make scholarly content open to the public, 82 per cent of 13 lakh articles published in Indian journals in a year are not publicly accessible (Openaccesindia, 2016). It has been over a decade and a half since the Budapest Open Access Initiative and only a handful of development researchers and organisations have taken steps towards making research open in India. To overcome this, in 2012, the Indian government released the National Data Sharing and Accessibility Policy enabling access to government-owned shareable data and information. Over the past few years, a number of departments in the Indian government along with private research organisations have launched or voiced their commitments towards open access in India.

Along with a growing push for transparent research practices and open access, there is a growing demand to strengthen what constitutes ethical research. Ethics form an integral part of transparency in research by defining standards that need to be upheld.

It is in this context, that 3ie and Sehgal Foundation organised an event for policymakers, government representatives, researchers and students to discuss transparency, open data and ethical values and issues related to development research.

Key highlights from sessions

Session 1: Transparency and open access in research
Panellists: Arul George Scaria, co-director, Centre for Innovation, Intellectual Property and Competition, National Law University; Avani Kapur, director, Accountability Initiative, Centre for Policy Research; Prabhakar Singh, executive director, Centre for International Legal Studies, O.P Jindal Global University; and Saurabh Bhajibhakare, senior research manager, J-PAL South Asia

Chair: Marie Gaarder, director, evaluation office and global director, innovation and country engagement, 3ie

Rapporteurs: Ritwik Sarkar, 3ie, Malvya Chintakindi, Sehgal Foundation

Summary

Panellists discussed their views on how data could be made open and how the understanding of open data differs across various disciplines. Marie Gaarder set the tone by expressing that even if data is available, it is not in a form that is usable. This is a challenge in low-and middle-income countries where data-driven decision making can improve lives. The panellists shared some of the challenges faced in meeting data and research standards; and making it publicly available, including getting stakeholders from NGOs, private organisations and research agencies to agree to common data standards.

Detailed discussion

Arul George Scaria (National Law University) defined open data as data without restrictions and easy to use. According to him, it is essential to be aware of confidentiality of data, something that the research community in India often fails to adhere to. He said securing informed consent from survey respondents, should be a priority. He also shared examples that showed how researchers in India do not share data in an open access platform. While there has been a move towards sharing data, sources of funding and research methodologies, there is a dearth of open access repositories which can house these materials.

Avani Kapur (Center for Policy Research) explained that open data relates to open research practices, making meta-data available and ensuring journal articles adhere to research standards. She highlighted that most research organisations in India lack the technical know-how of making data open and her team was currently engaged in collecting primary data and making existing data user friendly. She stressed that India needs better quality data than higher volume of data, there needed to be more engagement with donors to improve quality and access to data. She also outlined the challenges to ensuring transparency in research practices in India. According to her, researchers still work in silos, making collation of information difficult. Researchers lack the awareness as well as funding to support transparency initiatives. She pointed out that there is reluctance among researchers to ask for funding to make data open. Prabhakar Singh (O.P Jindal Global University) provided some legal context to the discussions, saying that there are a number of laws that govern data in India. He mentioned that data collection in India did not require in-depth technical knowledge. He also acknowledged that India tries to follow global standards on data protection, but researchers struggle to keep up.

Saurabh Bhajibhakare (J-PAL South Asia) gave a few examples to explain the importance of open data, citing a study conducted by Nature reported that 70 per cent of the 1600 researchers interviewed were unable to reproduce studies from fellow researchers and 60 per cent were unable to reproduce their own work. On a positive note, he added that 9 out of 10 top economics journals have made data accessible before the actual publication of a paper. He also outlined some challenges in the transparency efforts, for instance, preparing data for publication. He added that to ensure transparency in research, organisations need institutional capacity. He praised the efforts of 3ie, Pratham and others, saying steps like preregistration, pre-analysis plans and so on are very useful.

Session 2: State of open data in India
Panellists: Doug Johnson, director, IDinsight; Guneet Narula, DataMeet; Rakesh Ranjan, senior consultant, Niti Aayog, Government of India; and Pramit Bhattacharya, data editor, Mint, New Delhi

Chair: Sudarshan Ramaswamy, executive director, Centre for Ethics, Law and Political Economy, Jindal School of Government & Public Policy

Rapporteurs: Ananta Seth, 3ie

Summary
Sudarshan Ramaswamy (Jindal School of Government & Public Policy) pointed out that open source data is inherently a political issue. While the Indian government has attempted to promote open access through initiatives such as data.gov.in, data is still guarded and inaccessible. Rakesh Ranjan (Niti Aayog, Government of India) spoke about the efforts to make data publically available. According to him, while there is a consensus on promoting open access of government data, there is a need to quality assure data being shared. Pramit Bhattacharya (*Mint,* Indian financial daily) talked about the increasingly important role of data in journalism, in response to the credibility crisis and the growing demand of evidence-backed stories from an educated reader base. Guneet Narula (DataMeet) stated that while the government has approved the National Data Sharing and Accessibility Policy, the quality of data renders it unusable. Doug Johnson (IDinsight) discussed how accountability among the top echelons of the government can ensure better standards of data collection and analysis. He also argued that rigorous data collection should be ensured by coming up with a solution-oriented rather than a punitive approach. This discussion was followed by an interesting Q&A session, with the audience raising issues such as a lack of standardised data on the government's data-sharing platform, innovative methods of data collection, as well as data quality assurance practices.

Detailed discussion
In his opening remarks, Rakesh Ranjan shared that Niti Aayog has made progress in managing data and there is a clear understanding that data should be made public. According to him, the data has to be both accessible and usable. There should be a systemic approach to data and collection efforts should not be duplicated. He argued that due to multiple sources of data, it is often difficult to base decisions on evidence informed by data. There is a need for standardised systems of uploading block and district-level data to state and national repositories. He underlined the importance of third-party validation of data to quality assure data collected in field. He also spoke about Niti Aayog's aspirational district fellowship, with fellows working in 112 identified districts that have been ranked as the most backward based on 49 indicators, such as health, nutrition, education, skill development, agricultural productivity, water management, and infrastructure and so on.

Pramit Bhattacharya discussed the role of data in journalism. A proliferation of database tools has made government data more accessible for those working with newspapers and media

houses. This is in response to a credibility crisis which has encouraged journalists to separate facts and analysis from opinion. Institutions such as DataMeet have made visualising data easier by providing graphs and charts online. In response to demands for fact-based stories, data journalism teams are being set up across the country. This process has been informed by a bottoms-up approach- with the demand emanating from the reader base rather than being elite-driven.

Guneet Narula shared that the government approved the National Data Sharing and Accessibility Policy in 2012. The objective of the policy was to increase the accessibility and ease sharing of non-sensitive data amongst the registered users and their availability for scientific, economic and social developmental purposes. According to him, the government is outperforming the development research sector in sharing data openly and ensuring transparency. However, the quality of data makes it unusable for performing analysis. For instance, open transit data of Delhi buses has been released by the Delhi government. While, this database adheres to a global standard for transit data and the general transit feed specification; the data is not under the open data license and includes a clause barring 'malicious use' of the data. Moreover, when a group of data scientists analysed the data, they found a bus which covers about 500 kilometres in 30 minutes. He provided this as an example to highlight the concerns researchers have regarding the reliability of data collected by the government.

Doug Johnson shared that while many datasets are released by the Indian government, only ten per cent of these are frequently used by researchers. This includes the hugely useful NSSO, NFHS and census data, among others. There needs to be accountability among the top echelons of the government for ensuring standards of data collection and analysis. He also argued that if data being collected is not reflective of the realities in field, rigorous data collection should be ensured by coming up with a solution-oriented rather than a punitive approach.

Session 3: Ethics in research lifecycle
Panellists: Anant Padmanabhan, fellow, Centre for Policy Research; Anupama Jha, regional consultant, Trace International; L Venkatachalam, professor, Madras Institute of Development Studies; and, Stephen Marks, professor, Department of Global Health and Population, Harvard T.H. Chan School of Public Health.

Chair: Neeta Goel, senior evaluation specialist, 3ie

Rapporteurs: Saurabh Sood, Sehgal Foundation; Zeba Siddiqui, 3ie

Summary
The session was based on the premise that it is the collective responsibility of research community to promote ethics when conducting research. Anant Padmanabhan (Centre for Policy Research) shared his concern over an increasing risk of surveillance of citizens using technological innovations like self-learning algorithms. Anupama Jha (Trace International) made a strong case for responsible conduct of research by giving examples of how beneficiaries are

often excluded from the research process. L Venkatachalam (Madras Institute of Development Studies) brought up the issue of validity and reliability in data available with government agencies and insisted on ethical practice of conducting research in Indian Universities that suffers from serious issues like plagiarism. Stephen Marks, (Harvard T.H. Chan School of Public Health) briefly described the multiple issues relating to the conduct of academic research during its life-cycle.

Detailed discussion

Neeta Goel (3ie) started the discussion by sharing an anecdote from the 1990s, wherein a US governor's attempt to make health insurance data (for state employees) open for public access backfired. A PhD student was able to retrieve data pertaining to the governor by combining data from voter's list and the publicly available data on medical histories.

She used this example to highlight some of the risks around reusing publicly available datasets and the associated challenges related to accountability of reusing data. Despite taking the necessary precautions of removing the identifiers from the data before making it open access, the risks of it being combined with other datasets cannot be dismissed. She led panellists into a discussion on prioritising ethical considerations in research.

Anant Padmanabhan stated that ethics is not the top most priority among researchers in India. With the availability of massive datasets and technological advances (deep learning algorithms), there is a significant risk of surveillance, such as that allows profiling of citizens He differentiated the type of surveillance that private companies like Google and Facebook engage in from what the state can do with all the information it collects on its citizens, especially because the lay offers privacy only in an instrumental sense. According to him, the law offers privacy only in an instrumental sense. While there are laws to curb profiling risks and several compliance standards exist, it is also up to the citizens to know about their right to privacy. He mentioned, even though big data is gaining significance in the development sector, it is unable to overcome the inherent biases such as underrepresentation in datasets.

Anupama Jha, work works on anti-corruption initiatives in said that accountability and responsibility go hand in hand with transparency. According to her, we lack good intentions in the use of data, along with the issue of lack of inclusion of communities who are involved in the research. At times, vulnerable communities are used for drug trials without being informed or consented. She said that the solution to curbing ethical misconduct lies in training on ethics, across developed and developing countries.

L Venkatachalam pointed out issues of reliability and validity of data available with the government agencies. He talked about his experience during his master's course, where he had to collate data related to industrial pollution and crop productivity. After repeated attempts to reach out to multiple officials, he was only able to obtain data that was projected through arbitrary methods of estimation and did not hold any validity. He affirmed that at the village-level, information on basic indicators is not available. As an example, he pointed to the delay in producing damage assessment report during the recent cyclone in Tamil Nadu. Another

significant issue in data availability is sharing of data among agencies. In his experience, he has seen data being collected on indicators for which data was already available.

He also brought up how widespread the issue of plagiarism is other such issues in Indian universities. He explained how PhD students can hire 'ghost writers, to write and defend their thesis. He also gave an example where a student threatened a professor with a lawsuit, after the professor established that his thesis had 65% plagiarized content. He advocated for ensuring that current research methodology courses should look beyond plagiarism and include other aspects that ensure valid and reliable generation of data.

Stephen Marks talked about the ethical issues faced in across each phase of the research life-cycle. In the design phase, academic freedom and curiosity-driven choice of design may be compromised due to pressures of publications and career expectations. In certain cases, he said there is a conflict of interest when seeking funding. For example, cigarette companies funding projects to create a smoke-free world. He also pointed out that the process of obtaining consent may be culturally irrelevant for vulnerable communities which is an inherent flaw in the design structure. To ensure the research community is ethical is its conduct, he provided a few solutions. First, he called for ensuring vulnerable populations have agency, which is possible through participation in research. Second, he suggested adhering to regulations like the like Council for International Organizations of Medical Sciences (CIOMS) guideline/ Helsinki declaration that encourage the protection of human subjects in research. He also added that ethics and human rights are part of the Sustainable Development Agenda.