UNITED NATIONS ECONOMIC COMMISSION FOR EUROPE

CONFERENCE OF EUROPEAN STATISTICIANS

**Expert Meeting on Statistical Data Confidentiality**

26-28 September 2023, Wiesbaden

# Research data governance in low- and middle-income countries – Post-conference report

Pedro Ferrer Breda (Institution, Country or Organisation)

Pedro.ferrerbreda@uwe.ac.uk

Pedro Ferrer Breda, Felix Ritchie, Elizabeth Green, Cara Kendal.

University of the West of England, UK.

Research and policy development on the governance of confidential research data is dominated by the work of academics and government agencies based in high-income countries (HICs). This leaves three quarters of the world's population faced with a corpus of theory and good practiced guidelines which, although robust and well-established, makes little or no reference to the specific circumstances of low- and middle-income countries (LMICs). It may be that the data governance models developed in LMICS may be easily transferable to other contexts (there is some evidence, for example, that human-centred training adapts well), but in general there is little or no examination of this issue. There is however a large demand; a recent announcement of a training course in data governance for LMICS was 10x over-subscribed within the first two weeks of launch.

This paper reports on a virtual workshop held at the end of August 2023 which explored this topic with researchers and statistical organisations across LMICs. The topics included:

- How do attitudes to research data governance vary across countries?
- Are there key gaps in education, training, or information resources?
- Are there general lessons that can be applied across cultures and continents?
- Where and how do we develop capacity?
- Are there specific issues relating to LMICs which are poorly covered in HIC models (e.g. treatment of indigenous peoples)?

# Workshop purpose, structure and preparation

In 2001 a virtual conference on the present and future of microdata access[1] was organised by the same team at the University of the West of England Bristol (UWE). This identified areas of good practice

[1] Green, E., Ritchie, F., Tava, F., Ashford, W., & Ferrer Breda, P. (2021, July). The present and future of confidential microdata access: Post-workshop report. https://uwe-repository.worktribe.com/output/8175728/

and common agreement and noted that knowledge exchange between experts has been a significant driver of improvements in data governance.

However, both the conference and the authors' experience showed is a significant gap in knowledge about research data governance in LMICs. This workshop was designed to being addressing that imbalance. The discussion aimed to consider:

- What are the attitudes towards data sharing in LMICs?
- What are the challenges in using confidential data in LMICs?
- What is used as a guide for the governance of confidential data in LMICs?
- What happened in LMICs during the COVID-19 pandemic?
- How can we develop support networks to help countries with data governance training and the development of relevant models?
- How sustainable are current practices in LMICs?
- How can we develop consistent terminology without enforcing HIC cultural models?

Each of the six sessions (challenges and attitudes; needs and opportunities; next steps; all repeated for Easter and Western hemispheres) was split into two semi-structured discussions, using Google Jamboard™ to facilitate the conversation. The groups then reconvened to present and discuss findings. After the completion of sessions on Day 1 and Day 2, a report on the day's sessions was prepared by the UWE team of the key points and circulated to participants to stimulate discussion and reflection. A full report including appendices and further reflections will be made available on the University of the West of England Website.

# Reflections

In this section, we reflect on the participants' contributions and across all three days by themes. First, we consider what we have learned about the current state of play; second, we consider what are the key needs going forward, and what steps are needed to make some of this happen; third we explore the roles the various stakeholders play.

## Establishing the current state of play of data governance in LMICs

### Technology

**Data availability and quality** is a significant challenge in most LMICs. Data availability is restricted by resource limitations, and maintaining quality while preserving confidentiality is a significant challenge due to a lack of resources and knowledge. COVID-19 increased data availability, but maintaining accuracy, quality and privacy remained a challenge. Additionally, the absence of clear data architecture and common standards regarding metadata complicates access and use of existing data. While data may exist, researchers may not be aware of it due to confusing data architecture. Furthermore, the lack or inconsistency of metadata complicates effective data utilisation.

**Low bandwidth** in many countries complicates the use of adequate sharing, cybersecurity and analysis software. This leads to less safe ways of sharing data such as email. Inadequate cybersecurity further compromises safety of data shared through such methods. Additionally, the inability to use advanced analysis software can limit the usefulness of providing access to data. Participants also noted that significant differences in bandwidth between regions are common, particularly when comparing urban and rural areas. This complicates sharing data collected in faraway rural areas with academics and researchers typically concentrated in cities. During COVID-19, many organisations faced operational halts as a result.

Due to low bandwidth (and to lack of training in use of software), the inability of using some software has led to researchers continuing to use **paper-based data collection and physical storage.**

Generally, physical protection of data is carried to good standards; this extends to tablets, tape recorders and other etc. There are however concerns in cases where personal devices such as smartphones are used for data collection and storage. The adoption of digital collection and management of data is rapidly increasing, in part due to the pandemic.

Some participants were familiar **with research data centres/data enclaves**, but few had used them before. Only participants from Latin America had functioning microdata laboratories in their countries and had knowledge of this practice in a handful of other countries in the region. Based on this session, this could be limited to LATAM in LMICs, and Mexico appears to have had a role in the development of laboratories in the region. Other participants who had knowledge of similar infrastructure stated that data collected in their country was commonly stored and accessed in enclaves based in HICs.

The spread of **remote access to data** has increased data sharing, particularly since the COVID-19 pandemic. Some institutions are developing safe remote access using biometrics checks. However, pressure to reduce costs and speed up development makes it harder to ensure systems are set up correctly.

Most participants are unfamiliar **with statistical disclosure control and privacy-enhancing technologies**. The only region in which this workshop found the widespread knowledge and use of these technologies is Latin America.

### Organisation:

Both discussions and survey results suggest that the limited availability (and adequacy) of **guidelines and training** for data governance is a major issue in safe data access for research in LMICs. Most participants expressed concerns relating to insufficient knowledge and skills in data protection, data literacy and understanding of confidentiality in their organisations. In many cases participants also noted that they need better training themselves. There are limited opportunities to obtain training in LMICs, and difficulty in obtaining visas or funding for travel prevents many researchers from LMICs from presenting their work and receiving training in HICs. While the move online (partly because of COVID-19) has improved access to training, low broadband is a significant barrier in the delivery of interactive training. Furthermore, participants noted the lack in capacity of current training programmes (such as workshops, online courses) in data governance for researchers.

Many participants noted that **training designed in HICs is sometimes inadequate** in LMIC contexts. This disconnect between needs and practice limits researchers and owners' ability to make effective decisions when faced with LMIC specific factors which HIC training and guidelines do not account for. Training should account for laws, policies, ethics, and cultural norms of the context researchers engage in to meet the standards of communities of interest.

Training in analysis and confidentiality may be available, though it is **not always delivered to enough individuals involved in a project** due to funding constraints. Participants noted that while researchers may be trained, other researchers and crucial decision-makers sometimes lack data literacy and the understanding of data governance and confidentiality in general. Significant differences in knowledge increases risks of disclosure when sharing data with less trained researchers involved in the same project and complicates discussions around data sharing for research use.

Participants noted that project outcomes need to be clearly stated to gain approval for data access in their institutions. In general, processes were good in terms of matching the right level of detail in data to needs of research projects. Dissemination plan for results was often a requirement in approval processes. In some institutions, researchers cannot request for a larger scope after approval. They must start a different approval process from scratch if they wish to increase their scope.

Frequently, reliance on HIC funding leads to the **use of approval processes designed in HICs** to obtain funds. Participants noted that expectations and requirements of HIC approval processes often differ substantially from processes designed in LMICs. Where HIC institutions provide funding, LMIC researchers and communities are sometimes excluded (or less affected) from project benefits. For instance, many participants noted their names were commonly excluded from articles published by HIC organisations which they had taken part in. Lastly, the exclusion of LMIC individuals from the design of approval processes may prevent skill development in this aspect, which hinders any improvements in related issues described above.

**Low funding and short-termism** in funding limits the build-up of infrastructure, training capacity and maintaining staff post-projects in LMICs. Unequal funding, priorities and access to resources among institutions limits cross-organisational research. While the value of professional data managers is acknowledged, projects with less funding do not prioritise this.

Reliance on HIC funding complicates long-term capacity building for data governance. Additionally, **data is frequently managed in HICs** due to the lack of servers or adequate infrastructure to store and manage data in LMICs. This leads to a lack of control of LMIC data by LMIC researchers. Participants noted concerns about data colonisation by HICs institutions.

### Societal context:

**Data sharing competes with other priorities**, leading to limited attention and awareness. Adherence to regulations required for approval is commonly prioritised, though there is less concern about implementation of approved processes. Sometimes there are significant issues in terms of the ability to follow through ethical guidelines outlined at the start of a project. Privacy issues notably receive inadequate attention during health crises. This was most significant during COVID-19, as many expedited ethics approvals were granted.

The **limited communication and collaboration** are a significant hindrance to data sharing in LMICs. This is commonly both a cause and consequence of the lack of trust between organisations. Additionally, specialised individuals/departments within organisations typically don't communicate often. COVID-19 further reduced collaboration and isolated organisations, and in some cases this situation has not reverted.

**Establishing trust between and even within organisations in LMICs is challenging**. There is limited understanding and sensitisation of the consequences of sharing data, and in many LMICs cultural norms can promote data sharing without the regard to safety. Additionally, standards, priorities and understanding of "public good" generally vary between sectors (e.g., healthcare vs economics research). Therefore, participants expressed concerns that researchers (especially those not involved in data collection) may have limited understanding or care for privacy issues. In absence of formal elements in approval processes that ensure researchers' knowledge and adherence of data governance standards, trust is often based on personal networks.

Participants noted a **need for sensitisation of governments and key institutions to the benefits of data access**. Despite people becoming more aware of the importance of data access in crisis management during COVID-19, there are still significant issues in translating data driven insights for policy uptake. This limits the benefits data owners may expect from sharing data.

Additionally, lack of awareness in this respect leads to a lack of **involvement of governments in designing and improving policies and regulations** related to data governance. This has resulted in a lack of clear data management policies and inconsistent regulations, which participants note contributes to legal compliance concerns. Policies are in early stages of implementation, and HIC dominance significantly influences policy design, leaving grey areas in data protection for LMICs.

The **lack of awareness and engagement of the public also limits support** for data governance initiatives in LMICs. Concerns of monitoring during COVID-19 increased public interest in data sharing laws, privacy rights and data usage extent. This has led to progress in data governance in some LMICs; some participants noted that their country considerably revised (and mostly improved) data protection laws during or following the pandemic.

## Ways forward

### Improving training and information resources

The need for development and access to information resources and training has been a recurring topic throughout all the sessions. While this is likely in part because participants were largely self-selected (from a network of contacts including participants on the DRAGoN's summer school on data governance for LMICs), concerns about the absence/inadequacy of guidelines and training were shared by participants providing access to data or involved with data access policy in NSIs and other government organisations. Knowledge on technical topics was varied among respondents. For instance, few participants were aware of SDC. Many participants also noted that that available guidance was not always suitable to expectations set by their context. Unfortunately, following this event, we are still unaware of the full extent of the adequacy of HIC guidelines and training in the context of LMICs.

On the one hand, comprehensive technical guidance has already been developed in HICs and has been found to be easily transferrable to LMICs in certain contexts. Written guidelines are useful as templates. Therefore, better technical know-how requires more circulation of information materials, which can be achieved through knowledge exchange networks and more open-source publishing. However, DRAGoN's experience is that just providing written materials is of limited value, and training is needed first to help understand guidelines. Face-to-face training (online or in person) has been shown to be much more effective than passive forms of training in this regard, and live discussions with learners also help cater technical guidance and training to their specific context (interestingly, the community/user focused training in data governance developed by the DRAGoN team in HICs has transferred smoothly into a variety of LMIC contexts; the DRAGoN team will be presenting a separate paper on this in Spring 2024).

> ***Identified need: accessible guidance across the range of data governance activities, including templates.***

Even when technical guidance is available and researchers are aware of what needs to be done, the lack of funding, institutional support or just simple awareness may limit options in decision-making around data access. This highlights the **need for principles-based guidelines** which help frame strategies and decisions from the perspective of outcomes and goals rather than specific local conditions. For example, two ethics committees may find it hard to agree on the specific form of their approval process and questions. However, they can agree on what the purpose of the ethics committee is, what the approval process should cover. This can be the basis for delegation of authority for projects based in two institutions, such as a HIC funder and an LMIC partner.

Introductory principles-based guidance across many parts of the data governance framework has been developed and are widely used in HICs, and generally has good transferability. For example, the Five Safes data governance framework is increasingly used as a common frame of reference between and within countries, and the basis for more detailed discussions. The UK is currently developing a reframing of much of its data service governance using principles-based strategic planning. Increasingly this is feeding into legislation: the European GDPR, the UK Digital Economy Act and the Australian Data Access and Transparency Act are all principles-based in their research data governance.

***Identified need: high-level principles-based guidelines on data governance***

There is a concomitant need to **ensure that principles reflect the needs and interests of LMICs**, rather than simply adopting models used in HICs. These may be appropriate, but we don't have sufficient evidence to support this. Therefore, LMIC input into developing good data governance principles is important.

***Identified need: LMIC input for workshops developing good data governance principles***

Principles based guidelines are high level and less specific/prescriptive. Therefore, **practical guides for implementation relevant to specific contexts** are needed to bridge the gap between theory and practice. This requires the study of practical examples from organisations relevant to LMICs when developing guidelines. These should be linked to criteria, pros and cons, risks and constraints which may be specific to some LMICs. The difference in the level of knowledge, access to infrastructure and institutional backgrounds described by participants demonstrates the need for further consideration of local contexts before attempting to develop guidelines based on the simple typology of LMIC vs HIC. This requires the involvement of local stakeholders in the co-development of good practice guidelines and case study research by academia.

***Identified need: specific case studies and examples reflecting the likely needs of LMIC users***

In cases where local institutions and communities of interest have limited or no experience in data governance, the role of knowledge exchange networks is important in providing examples of precedents in similar contexts. National or regional champions are key in the development of such networks, directly aiding the development of guidelines and providing training. Identification and support of champions may require involvement of international organisations and academia.

One way to bring together effectively written materials, high-level and detailed guidelines, face-to-face and passive learning, is to build a hierarchical training model. Attention is focused on providing relevant, high quality, perhaps resource-intensive training to selected individuals, who will then take that training to:

- Develop additional, locally relevant materials.
- Interpret generic guidelines for local audiences.
- Train (or commission/support the training of) further individuals, in a snowball effect.

***Identified need: develop the tools and resources to support hierarchical training models including training for champions and train-the-trainer materials***

### Accreditation and auditing:

Participants proposed the development and adoption of **formal accreditation systems** as a solution for ensuring safe data sharing and improving trust. For example, existing approval processes in LMICs generally only check the identity of researchers to determine their trustworthiness. This is not sufficient to demonstrate knowledge and adherence of data governance practices and is only useful as basis of threats for legal consequences of disclosure. Personal networks are sometimes used as the basis for data sharing where formal processes are absent or perceived as inadequate, though this does not allow sharing outside of such existing networks and is especially an issue where data from multiple departments/organisations is needed.

Formal accreditation systems also **provide a link between principles-based planning and implementation**. Principles-based planning is an efficient, user centred way to address strategic problems in data governance, but by its nature it does not specific how implementations are to be done. Accreditation is the link. In a principles-based system, a good accreditation system seeks to ensure that a specific implementation is aligned with and satisfies the strategic principles; it does not by prescribing what must be done, but by measuring whether the proposed solution meets the strategic goals. In several HICs, accreditation is increasingly based around the Five Safes: for example, identifying what a 'safe researcher' or a 'safe system' is independent of any specific project.

Ideally, accreditation processes **have validity outside their local context**: an accredited ethical review process recognized by others can be the basis for a data sharing partnership. In the UK, users of Trusted Research Environments (TREs) all go through similar (but not identical) training and vetting processes. The result is that all TREs accept the 'safe researcher' status conferred by others, and researchers are free to move between TREs without further accreditation.

***Identified need: models of accreditation that can be adapted and adopted***

## Changes in attitudes:

Many participants reported **significant attitudinal barriers** to effective data sharing, particularly in cases where researchers had access to relatively better training and infrastructure. These attitudinal barriers may exist commonly in LMICs but are generally found to be more significant after other more observable barriers, such as the lack of know-how and resources are overcome. Nevertheless, a change in attitudes is often key in enabling the improvement of capacity for data governance. Based on this workshop, we identified the attitudes of data owners, data users, government institutions, funding agencies and the public as particularly important in enabling better data sharing. **Data owners' and users' attitudes** may be changed through the delivery of training and the development of accreditation systems. Better knowledge of options, benefits and costs in decisions around data access can, all other things equal, lead to a change in data providers/owners attitudes. Moreover, accreditation systems help build trust between data owners and decision makers. This allows data owners to encourage/promote good practice. Auditing can then help adhere to standards and transparency.

***Identified need: support the development and adoption of accreditation and auditing to change the attitudes of data owners and users.***

There is a **need for sensitising governments to the importance of data governance**. Doing this will help the development of clear legal and regulatory frameworks for data sharing. In many cases, frameworks are unclear or make no reference to data sharing for research. Additionally, sensitising governments provides institutional consistency and institutional backing for good governance and privacy protection. Involvement by governments in policy design may **discourage the perception of data as a produc**t. For example, where funders consider that they "own" data created as part of a project.

***Identified need: engage with government agencies to sensitise them to the importance of data governance.***

**Building community trust and awareness** is becoming increasingly important, especially following COVID-19. Mobile technology is growing in LMICs and is being used increasingly for data collection. There is a need to reassure communities about the security of data collected on these devices. General increased trust in research use of data is therefore needed to address this. Participants

suggested LMICs typically have high engagement with communities at data collection level. This offers opportunities to build trust which might be missed by HIC researchers.

> ***Identified need: explore and document examples of building good trust at data collection level***

### Knowledge exchange networks

Based on this workshop and DRAGoN's experience, there is little communication between LMIC experts, except for some countries in Latin America. Even in HICs, there is a small pool of experts in policy design for the governance of confidential data for research use. As a result, knowledge exchange networks have proved essential in sharing experience, good practices and guidelines. In LMICs, the number of experts is even smaller, and in some cases non-existent.

Many **institutions in HICs who engage with LMIC agencies** have an important role in including LMICs in networks and aid the development of new networks. For example, DRAGoN works closely with LMIC researchers through its data governance training courses to develop common practices and understanding in LMICs.

> ***Identified need: HIC agencies should support networking between LMICs.***

Some LMIC agencies have already adapted or developed guidelines and have the capacity to support the implementation of good data governance practices in similar LMICs. This is the case of **regional champions** such as Mexico's INEGI, who are supporting the adoption of microdata laboratories in a few Latin American countries.

> ***Identified need: Identify and support potential regional champions.***

Lastly, **engagement with non-obvious groups can be used for network building**. For instance, events such as UN-ECLAC's statistics conference of the Americas may not include participants who are experts in confidentiality. However, individuals from such agencies are key in the sensitisation of agencies to the importance of data governance and can act as a bridge with individuals responsible for data governance in their agency.

## Who can do what?

| Who? | What? |
|---|---|
| International agencies and NGOs | - Practical experience<br>- Offer advice/examples of good practice<br>- Capacity building<br>- Enforcement of standards<br>- Make governance explicit in funding |
| NSIs, local research institutions and public health bodies in LMICs | - Practical experience<br>- Develop institutional policies and guidelines<br>- Experience for adjusting to local circumstances |

| | |
|---|---|
| National and regional champions | - Practical experience<br>- Advice/examples of good practice<br>- Capacity building<br>- Enforcement of standards<br>- Help understand value and risks of what is being done<br>- Case studies/local examples |
| IRBs and other actors | - Develop institutional policies and guidelines |
| Professional practice and research associations | - Enforcing standards<br>- Supporting development of accreditation |
| Academia | - Case studies/local examples<br>- Designing principles and guidance<br>- Offer advice/examples of good practice |

Many stakeholders have a role in designing and enforcing standards. However, there can be some conflict between who decides standards and who enforces them.

At this stage it is not clear how a coalition can be built to address these issues. However, we want to encourage discussions, and we will form an informal group in data governance. We aim hold an initial meeting in winter 2023-2024. We wish to identify people who might want to contribute or support this. If you would like to join us, please contact:

- Dragon@uwe.ac.uk

To contact the authors of this report directly:


- Pedro.ferrerbreda@uwe.ac.uk
- Felix.ritchie@uwe.ac.uk
- Elizabeth7.green@uwe.ac.uk



FUNDED BY NIHR National Institute for Health and Care Research

UKaid from the British people

UWE Bristol University of the West of England

Data Research Access & Governance Network