Research data governance in low- and middle-income countries

Post-workshop report

October 2023, Bristol

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

Data Research Access and Governance Network (DRAGoN dragon@uwe.ac.uk)

University of the West of England, Bristol

Corresponding author: Pedro2.Ferrerbreda@live.uwe.ac.uk

Contents

Executive summary

'Data governance' concerns all aspects of the effective and efficient use of data, including data management, ethics, and disclosure control. Data governance is particularly important for the safe use of confidential data. Data governance for research studies is largely the province of high-income countries (HICs), where almost all of the theory, guidelines, tools and understanding of good practice were developed.

In August 2023 a week-long virtual workshop was held on research data governance in low and middle-income countries (LMICs). Separate sessions were held to allow participation from Eastern and Western hemispheres. This workshop explored the current state of play, challenges and opportunities in the governance of confidential data for research in LMICs. It also aimed to facilitate communication and build networks which may help to address problems identified. Some 40 participants attended, mainly from Africa, Latin America, and West and South Asia. Participants were sent a pre-workshop survey to help frame workshop discussions.

The workshops explored three topics over the week: attitudes, main challenges and what happened during COVID; what is working well and what isn't; ways forward for data governance in LMICs.

The **current state of play** was summarised under three headings:

Technology

- Data availability and quality is a challenge.
- Low bandwidth is common.
- Paper-based data collection and physical storage.
- Research data centres/data enclaves are rare.
- Remote access to data has increased, but there are issues.
- Most researchers are unfamiliar with SDCs and PETs.

Organisation

- Guidelines and training are needed.
- Training designed in HICs is sometimes less adequate in LMICs.
- Approval processes are commonly designed in HICs.
- Low funding and short-termism limits long-term capacity building.
- LMIC data is sometimes managed in HICs.

Societal context

- Data sharing competes with other priorities.
- Limited communication and collaboration between agencies.
- Low trust between (and within) organisations.
- Governments and key institutions are often unaware of importance of data access.
- Governments are less involved in data governance policy.
- Less public awareness and engagement limits support for initiatives.

In terms of **ways forward**, the workshop identified the need for

Training and information

- Accessible guidance across the range of data governance activities, including templates.
- High-level principles-based guidelines on data governance
- LMIC input for workshops developing good data governance principles.
- Specific case studies and examples reflecting the likely needs of LMIC users.

- Support for hierarchical training models including training for champions and train-the-trainer materials.

Accreditation and auditing

- Accreditation provides a link between principles-based planning and implementation.
- There is a need for models of accreditation that can be adapted and adopted in LMICs
- Accreditation could be based on the dominant HIC governance framework, the Five Safes
- Should have validity outside of local contexts

The workshop identified the need for change **in the way actors approach** data governance:

- Data owners' attitudes
- Data users' attitudes
- Government institutions' attitudes
- Funding agencies' attitudes
- Public attitudes and engagement
- Knowledge exchange networks

It was suggested that there **a substantial shortage of expertise** in data governance policy design in LMICs; to ameliorate this, LMICs should be included in existing networks, and LMIC specific networks should be developed. Regional, national and sectoral champions have a key role here.

Finally, the workshop considered how to make change happen: **who can do what**? These stakeholders who may have a role in supporting good data governance:

Who?	What?
International agencies and NGOs	• Practical experience • Offer advice/examples of good practice • Capacity building • Enforcement of standards • Make governance explicit in funding
National statistical institutes, local research institutions and public health bodies	• Practical experience • Develop institutional policies and guidelines • Experience for adjusting to local circumstances
National and regional champions	• Practical experience • Advice/examples of good practice • Capacity building • Help understand value and risks of what is being done • Case studies/local examples
Institutional/Ethical Review Boards, and related actors	• Develop institutional policies and guidelines
Professional practice and research associations	• Enforcing standards • Supporting development of accreditation
Academia	• Case studies/local examples • Designing principles and guidance • Offer advice/examples of good practice

1. Workshop purpose, structure, and preparation

1.1 Aims and objectives.

'Data governance' concerns all aspects of the effective and efficient use of data, including data management, ethics, and disclosure control. Data governance is particularly important for the safe use of confidential data.

In 2021 a virtual conference on the present and future of microdata access[1] was organised by the team at the University of the West of England Bristol (UWE). This identified areas of good practice 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 Eastern 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. These daily reports are included in the Appendices, and are summarized in the main body of this text.

1.2 Pre-conference survey

Prior to the conference, participants were invited to complete a survey to inform the preparation of discussions. This survey asked general questions about participants' role and experiences regarding data access for research. Tables 1 and 2 provide information on the characteristics of participants who responded to the survey.

Data type	Yes	No	Percentage yes
Economic	5	17	22.73%
Environment	1	18	18.18%
Health	20	2	90.91%
Social	6	16	27.27%
Other	1	21	4.55%

Table 1 Findings from pre-workshop survey: types of data being used by participants

[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/

Organisation type	Number	Percentage
Government or Public sector	9	42.86%
Health	3	14.29%
Private Sector	0	0%
NGO	4	19.05%
Other	5	23.81%

Table 2 Findings from pre-workshop survey: sector breakdown of participants

Two thirds of the 31 respondents are data users. The remaining third are involved in providing/supporting access for research use or involved in research for data access policy. Most work in governments/public sector, notably in public healthcare. This is reflected in the fact that 22 participants work primarily with health data. Some also work for NGOs (6) and academia (6), and a few participants also use social (6), economic (5) and environmental (4) data.

Respondents are familiar with ethical review boards, and stated they are familiar with anonymisation (however, subsequent discussions at the workshop showed very different interpretations of 'anonymous'). But beyond this, respondents had less familiarity with other data governance concepts such as the Five Safes, and almost none on technical matters such as statistical disclosure control.

Overall, stances towards data access are perceived to be default-closed (see Glossary). On average, 86% of researchers rate their organisation's stance as closed, and 84% their own stance being closed. Survey respondents' personal views in this matter often match their organisation's perspectives. This is unusual: similar surveys in HICs tend to show that individuals are default-open personally but believe their organisations are default-closed.

Half of respondents replied that their organisations require researchers to undergo training in data governance. The rest stated that this is not required, or are unaware of whether training is required. The objective of training is often geared towards ensuring compliance with rules during research; training in managing risk form outputs is rare[2]. Only two participants use secure research facilities.

Participants were then asked to provide additional comments on challenges around data governance and potential solutions to these challenges. All of the few responses from these questions highlighted issues that came up in the workshop sessions, and are not repeated here.

2. Session summaries

2.1 Context, attitudes, barriers, and COVID-19:

This session discussed the social and institutional context, attitudes, and general barriers surrounding data access for research use in LMICs. It also reviewed participants' perspectives on what happened in this regard during COVID-19, summarised below. **Appendix 1** contains a detailed report of this session.

1. Data sharing and governance often compete with other priorities. While the importance of good governance is growing, in part due to the influence of funding agencies, established norms commonly promote data sharing without regard to safety. COVID-19 reinforced this stance.

[2] Best practice in training is seen as focusing on community building rather than the explanation of rules; and training in statistical disclosure control is usually considered important. Green et al. (2021)

2. Maintaining data quality while preserving confidentiality is a challenge. During COVID, data availability surged, but maintaining accuracy and confidentiality remained a challenge.
3. Establishing trust between organisations is challenging due to fears of misuse and misinterpretation of data. Limited resources and training and lack of communication between organisations greatly contribute to this issue. This poses challenges for collaboration and access by external users.
4. There are significant gaps in skills and training for users and owners.
5. There is a lack of clear data management policies and governance structures. Coupled with limited guideline documentation, this leads to inconsistency, legal compliance concerns and confusion. Approaches also vary with data types (routine use vs project specific data).
6. Policies are in the early stages of implementation and are frequently influenced by HICs. This leaves grey areas in data protection for LMICs.
7. Limited infrastructure hinders secure data storage and sharing, leading to insecure practices. Concerns were raised about solutions requiring third parties, often in HICs. This is can arise as a short-term expedient, since external funding is often short term, but this hinders local infrastructure development. Furthermore, this creates further challenges in data access and control arises when servers are located in HICs.
8. The requirements of HIC funding may have influenced perceptions of data as a monetary product and raise concerns around the lack of control of LMICs of their LMIC data and data colonisation.
9. Public trust was strained over COVID due to concerns over monitoring. This crisis also saw an increase in public interest in data sharing laws, privacy rights, and data usage extent.

2.2 Challenges and opportunities

This session comprised an in-depth discussion of challenges and opportunities around data access in LMICs, summarised below. **Appendix 2** contains a detailed report of this session.

1. There are clear differences between participants' experiences across all aspects of data governance, highlighting the potential inadequacy of one-size-fits guidelines for the governance of data for research in LMICs. Not only do standards differ across countries, but differences are evident within the same country based on sector and specific institution.
2. However, several participants' accounts may allow the formation of groups who share challenges in common. For instance, participants from countries which have relatively less strong research institutions expressed concerns regarding the lack of control on their data caused by the reliance on funding from HICs.
3. Several participants also agreed in their accounts of the treatment of data as a 'product' (that is, with same expectations of being able to own buy or sell data as for any other product) in their context, often in cases where research was funded by HICs.
4. While many participants described formal processes assessing the trustworthiness of data users, accounts differed on the importance given to users' training in data management and confidentiality in such processes.
5. While all participants agreed in the importance of confidentiality during research, accounts showed many interpretations of this. Furthermore, this agreement may be a result of the selection of participants, which was mostly through snowballing the UWE team's contacts as external assessors and deliverers of data governance training.

2.3 Ways forward

This session consisted of a discussion concerning potential solutions to challenges from sessions 1 and 2, including a consideration of the roles of different stakeholders in attaining such solutions.

Additionally, live polling was used to determine the relative importance given to these solutions by participants for points 1 and 2 below.

1. Most agree that data governance is a priority; around half of participants' projects involved a thorough consideration of data governance.
2. Participants believe that the largest improvement to data governance in LMICs would be a greater awareness of the topic among researchers, followed by greater availability of good practice guidelines, training of staff, a change in cultural attitudes to privacy, better infrastructure and better technical resources.
3. There is a need for information resources, namely principles-based introductory guidelines, practical guides for implementation, technical training and comprehensive guides linked to criteria, pros and cons, risks and constraints[3].
4. Developing information materials useful for LMICs should involve local LMIC stakeholders. Regional champions and knowledge exchange programs are very helpful in providing guidelines and precedents in similar contexts.
5. Participants highlighted the need for sensitising governments to data governance. This may help create institutional backing for good governance and privacy protection and institutional consistency.
6. A clearer legal and regulatory framework is needed.
7. Research into the value of long-term planning conducted by academia and supported by international institutions is needed to change attitudes in government.
8. Face-to-face (active) training is most effective, whether online or in person. Capacity building therefore relies on training key individuals to train other researchers. Additionally, there is a need for being able to prioritise the right individuals who need training to make the best use of existing capacity.
9. Formal accreditation systems are needed to demonstrate knowledge and adherence to standards. These could be developed by public health bodies, academia, NSIs, international agencies and NGOs, professional, practice and research associations.
10. Funding is still a huge issue for good data governance in LMICs. Participants noted a need for making data governance explicit in funding.

[3] Principles-based planning and design focuses on what is to be achieved and the approach to problem solving, and acknowledges that there may be different ways to achieve the outcomes. This is in contrast to rules-based guidance, where the aim is to be as specific as possible about necessary actions.

3. 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 could play.

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

3.1.1 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, voice recorders 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** (RDCs), but few had used them before. Only participants from Latin America had functioning general-purpose RDCs in their countries; they also 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.

3.1.2 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 from 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 from (or less able to realise) 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, hindering 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.

3.1.3 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** is 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 sometimes 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 the 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.

3.2 Ways forward

3.2.1 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 courses on data governance for LMIC health researchers), 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 statistical disclosure control (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 (for example, in the DRAGoN data governance courses online and delivered face-to-face in Nepal). 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, and 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 and the Five Safes. 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.

3.2.2 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.

3.2.3 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.

3.2.4 Knowledge exchange networks

Based on this workshop and DRAGoN's experience, there is limited communication between LMIC experts, with the exception of 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 in HICs have proved essential in sharing experience, good practices, and guidelines. In LMICs, where the number of experts is even smaller and in some cases non-existent, the value of such interactions is likely to be much higher.

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.

4. Next steps: who can do what?

Based on discussions and DRAGoN's experience, we have listed a number of stakeholders and possible roles they could have in improving the governance of confidential data for research in LMICs, and what each of these roles would involve.

- **International agencies and NGOs** who work directly with government, healthcare, and research institutions in LMICs can offer advice and examples of good practice drawing on their experience with similar situations from other organisations they work with. Additionally, they can provide funding for capacity building. They also have a role in making governance explicit in funding.
- **Local research institutions, public health bodies and NSIs** can leverage their experience in local circumstances to support the development of relevant institutional policies and guidelines.
- **National/regional champions** have a role in supporting data governance. Research, healthcare, and statistics institutions may contribute to the development of guidelines, policies and training in other LMICs with similar contexts. Mexican INEGI's and INSP's involvement in the development of microdata laboratories and training of researchers,

respectively, has shown to be effective so far. Additionally, connections made by regional champions in the development of regional knowledge exchange networks.
- Institutional Review Boards have a role in developing policies and guidelines.
- Professional practice and research associations should ensure standards are met. They have a role in supporting accreditation and auditing.
- **Academia** has a role in supporting the design of principles and guidance based on their experience as data users and in policy design. This involves the where possible, case study research.

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

4.1 Further information and contact details

At this stage it is not clear how a coalition can be built to address these issues. However, we want to begin discussions, and we will be forming 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

Appendix 1: Daily summary of session 1 (attitudes, awareness, current status)

What are the attitudes towards data sharing in LMIC's?

1. **Prioritization and Awareness Gap**: Data sharing competes with other community priorities, leading to limited attention and awareness.
2. **Privacy during health Crises**: Privacy issues receive inadequate attention during health crises, impacting public trust.
3. **Trust, Relationships, and Expertise**: Establishing trust between organisations is challenging due to limited resources and expertise, impacting data sharing success.
4. **Researcher Perspectives and Skills**: Researcher skill gaps hinder data sharing efforts, especially when compelled to share data without adequate skills.
5. **Guideline Diversity**: Inconsistent guidelines on organizational approaches, ownership, and protection create confusion.
6. **Infrastructure and Third Parties**: Limited resources necessitate third-party involvement, hindering local infrastructure development.
7. **Cultural Acceptance and Policy Lag:** Cultural norms can promote data sharing without regard to safety, policies are in the early stages of implementation.
8. **HIC Dominance**: HIC dominance influences policies, leaving grey areas in data protection for LMICs. Worries about data colonization.
9. **Contributor Impact and Reputation**: Data collectors are often sidelined, impacting their credit. Additionally, there are fear that data will be misused when in external control.
10. **Data Value and Protection**: Some value data monetarily, leading to reluctance in sharing outside organizations.
11. **Privacy Concerns and Mistrust**: Data misuse and misinterpretation fears create mistrust among individuals and organizations.
12. **Awareness Gap and Consequences**: Limited understanding of data's consequences hinders sharing with untrusted entities.
13. **Ethics, Consent, and Hesitancy:** Complexities of ethics and consent contribute to sharing hesitancy.
14. **Growing Data Sharing Importance:** Data sharing's importance grows due to funders' emphasis.
15. **Political Sensitivity**: Politically sensitive data is guarded due to risk of misuse, esp. health data.
16. **Complex Data Access**: Accessibility varies with data types (routine vs study specific), influencing sharing practices.

What are the Challenges of using confidential data in LMIC's?

1. **Resource Limitations and Insecurity:** Limited financial and technical resources hinder secure data storage and sharing, leading to insecure practices. Financial constraints hinder access to cloud storage platforms, causing data shortages. Reliance on paper-based resources.
2. **Balancing Data Quality and Confidentiality**: Maintaining data quality while preserving confidentiality is a challenge, particularly for non-RCT studies.
3. **Documentation Deficiency and Utilization:** Poor metadata and ethical clearance documentation complicate effective dataset utilization.
4. **Server Location Impact:** Difficulty in data access and control arises when servers are located in the global north.

5. **Policy and Governance Gaps:** Lack of clear data management policies and governance structures, coupled with limited guideline documentation, lead to inconsistency and legal compliance concerns. Absence of clear architecture, regulations, and documentation compromises data quality and security.
6. **Skills and Training Gap:** Insufficient training in data protection limits knowledge and skills. Insufficient data literacy and understanding of classifications complicate data management.
7. **Data Breach and Reputational Risk:** Inadequate cybersecurity, limited resources, and financial constraints lead to breaches and unauthorized access, impacting reputation and trust.
8. **Data Governance Team Absence:** The lack of a data governance committee adds complexity to data management.
9. **Institutional Disparities and Accessibility:** Unequal access to data and resources among institutions limits broader research. Public good focus varies.
10. **Collaboration Challenges:** Limited collaboration hampers data sharing.
11. **External Funding Limitations:** Short-term external funding doesn't sustainably improve infrastructure.

What happened in LMIC's during the COVID-19 pandemic?

1. **Publication and Data Access Shift:** Publication standards became more rigorous, demanding proper documentation. Data access varied, with some countries becoming more open and others tightening control due to digitalization.
2. **Open Access Emphasis:** Open access data gained focus.
3. **Remote Research and Operations:** Remote research and work methods gained traction, while some organizations increased data sharing and research activities, others faced operational halts.
4. **Government Control and Ethical Implications:** Governments exerted more data control, especially from government institutions. Expedited ethics approvals were granted for COVID-19 research, posing ethical challenges.
5. **Data Accuracy and Delays:** Data availability surged, but maintaining accuracy remained a challenge. Underprepared systems led to data delays, necessitating secondary sources.
6. **Trust and Interest in Data Laws:** Public trust strained due to concerns over monitoring. Increased public interest in data sharing laws, privacy rights, and data usage extent remained, shifting data perceptions.
7. **Technological Disparities and Digital Shift:** Uneven technology access caused remote work disparities. Traditional data transfer methods shifted to digital means, raising safety concerns.
8. **Collaboration and Training Challenges:** COVID-19 restrictions reduced collaboration and isolated organizations. Research and training moved online, shaping future practices.
9. **Adaptation in work:** Outdoor fieldwork adaptation emphasized hygiene and potentially impacted future practices. Limited fresh data collection led to reevaluating preexisting data applications and adopting new perspectives. Online work took over and VPN use grew.
10. **Data's Role Emphasis:** The pandemic emphasised data's vital role in crisis management, fostering open data attitudes.

Appendix 2:Daily summary of session 2 (challenges and opportunities)

Key points:

1. There are clear differences between participants' experiences across all aspects, further highlighting the potential inadequacy of one-size-fits guidelines for the governance of data for research in LMICs. Not only do standards differ across countries, but differences are evident within the same country based on sector and specific institution.
2. However, several participants' accounts may allow the formation of groups who share challenges in common. For instance, participants from countries which have relatively less strong research institutions expressed concerns regarding the lack of control on their data caused by the reliance on funding from HICs.
3. Several participants also agreed in their accounts of the treatment of data as a product in their context, often in cases where research was funded by HICs.
4. While many participants described formal processes assessing the trustworthiness of data users, accounts differed on the importance given to users' training in data management and confidentiality in such processes.
5. While all participants agreed in the importance of confidentiality during research, accounts showed many interpretations of this. Furthermore, this agreement may be a result of the sample's method of selection, which was mostly obtained through DRAGoN's contacts as external assessors and deliverers of data governance training.

Safe projects:

1. **Ethical review processes:**
    a. In some contexts, thorough processes are widely used, and rules are clearly defined.
    b. However, they are much less used in other countries and in some contexts, there are significant issues in terms of the ability to follow through with ethical guidelines outlined at the start of the project.
    c. Adherence to regulations required for approval is prioritised, though there is less concern about implementation of approved processes.
    d. This is sometimes the case even when undertaken by qualified individuals, and in instances may result from a lack of understanding of the importance of confidentiality.
    e. Researchers in some countries struggle with ethical processes.
    f. Sometimes HIC's write ethics applications for LMIC projects, preventing skill development.
2. **Approval processes:**
    a. Project outcomes need to be clearly stated in order to gain approval.
    b. In general, processes were good in terms of matching access to appropriate levels of detail in data to the needs of research projects.
    c. **Dissemination plan** for results was often a requirement in approval processes.
    d. In some institutions, researchers cannot request for a larger scope after approval. They must start from scratch if they wish to increase their scope.
3. **Funding:**
    a. Concerns were raised about having to use approval processes designed in HICs to obtain funding for projects in LMICs. Expectations and requirements for Principal Investigators in HICs and LMICs differ substantially in many cases.
    b. Where HIC institutions provide funding, LMIC researchers and communities are frequently excluded (or less affected) from project benefits.
    c. There are financial constraints on maintaining staff to share data after projects close.
    d. Projects funded by HICs were commonly focused on the output of projects and deliverables rather than the process for generation of outputs.

4. **Engagement:**
    a. Stronger community ties and social networks help with engagement throughout entire project lifecycle.
    b. Feedback to communities' data is collected from is uncommon.
    c. Engagement is difficult for short life projects to manage community engagement properly – this is harder when there are financial constraints.
5. **Post-project:** Data can be used in resource mobilisation to scale up interventions.
    a. Some institutions require data to be shared with up to 2 years post-project end. Some participants expressed concerns on their lack of control of data post-project.
6. **Community networks:** LMICs often have strong community networks. These can help with buy in and long-term use of project outputs.
7. **Academic writing and approval:** A lower level of academic writing skills (by HIC standards) sometimes limits ability to get funding, especially where projects rely on funding from HICs.
8. **Data sharing agreements:** role of PI or people responsible for data collection are not often considered in data sharing agreements.
9. **Issues with equity**
    a. Between and within countries due to socioeconomic and geographic factors
    b. Uneven provision of infrastructure in countries/regions.
10. **Issues with translating data for uptake by policy** which may impact communities' data is collected from.

Safe people:

1. **Training:**
    a. In general, larger projects are often funded to include training, smaller projects are not.
    b. In some countries training is a common requirement in the health sector.
    c. Good clinical practice training is often required for randomised controlled trials. This was mostly delivered online.
    d. Training should consider law, policy, ethics, and cultural norms of areas researchers engage with in order to meet the standards of affected stakeholders.
    e. Better training in analysis skills is required in some instances. This is constrained by funding.
2. **Data managers:** Participants generally agreed in the importance of data managers. However, in projects with less funding the hiring of data managers was often not a priority.
3. **Disconnect between needs and practice:** In instances training is widespread and satisfactory, but is not delivered to all individuals involved in research. Decision makers sometimes lack understanding of data governance and confidentiality in general.
4. **Equity issues due to mobility:** Difficulty in obtaining visas or funding for travel prevents many researchers from LMICs from presenting their work and receiving training in HICs.
5. **Trust:** There is a lack of confidence that researchers not involved in data collection may have limited understanding or care for privacy issues.
6. **Checks:** Identity checks are often required; references and qualifications are less common.
7. **Networks based on personal trust in data sharing:** personal networks are sometimes used as the basis for data sharing where formal processes are absent or potentially less adequate.

Safe settings:

1. **More experience in paper-based data:** Thus, physical protection of data is carried to good standards. However digital collection and management of data is rapidly improving, possibly brought on by the pandemic.

2. **Remote access:** Some countries are developing safe remote access using biometric checks. Pressure to reduce costs and speed up development makes it harder to ensure systems are set up correctly.
3. **Secure Access Facilities:** Some countries have functioning microdata laboratories. Based on this session, this could be limited to Latin America in LMICs. Many other LMICs store their data in secure access facilities based in HICs, which limits LMICs' control over their data.
4. **Lack of funding** and resulting inappropriate infrastructure can limit security.
5. **Analysis software:** Can be limited or unavailable. Lower internet connection can limit newer, more powerful software which requires more bandwidth.
6. **Sharing:** sharing data in secure ways through secure servers is common in many countries, while other participants noted that less secure sharing through email or personal use devices is also common.

Glossary

To aid discussions in this workshop, we have compiled a glossary based upon the 2021 conference (footnote). This document reflects an HIC perspective and is not intended to enforce any particular views and be a definitive or replace other glossaries.

Term	Definition to be used in the workshop
Anonymous data	Data that does not include sufficient detail to allow the data subject to be identified, under any reasonable conditions
Breach of confidentiality	The release of identified or de-identified data to an unauthorised system, environment, or person; a breach of confidentiality may not mean a disclosure as it will depend on the circumstances
Breach of procedure	Failure to follow appropriate operating procedures, irrespective of whether a breach of confidentiality occurs
Confidentialisation	The act of reducing the likelihood or harm of re-identification by reducing detail or perturbing the dataset
De-identified data	Data which includes sufficient detail to allow the data subject to be identified, but only with effort and with less certainty (for example, a combination of gender, age, type of employer, salary range and disability status)
Distributed access	Restricting the physical location of the data, but allowing users in other locations to carry out analysis and retain statistical results (but not microdata)
Distributed data	Sending microdata to users under licence, to analyse on their machines
FAIR principles [4]	An acronym for Findability, Accessibility, Interoperability, and Reusability
Five safes[5]	The Five Safes is a framework to aid decisions about the effective use of data that is confidential or sensitive. The Five Safes model also places statistical disclosure control (SDC) in its proper context, as part of a system approach to data security. The Five Safes breaks down the decisions surrounding data access and use into five related but separate dimensions: safe projects, safe people, safe data, safe settings, safe outputs.
Identified data	Some data directly (not necessarily uniquely) relates to an individual respondent e.g., name, detailed address, social security number, Health service number, tax registration number etc
Input SDC	The application of SDC methods to raw data to reduce data risk before it is released to the users
Microdata	The individual unit records about a person or organisation, such as information collected from surveys or administrative data

[4] **Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., ... & Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3: 160018.**

[5] **http://www.fivesafes.org/**

Remote access	A system that allows users to 'see' and manipulate the source data
Noise:	Noise refers to a random alteration of data/values in a dataset so that the true data points (e.g., personal identifiers) are not as easy to identify.
Output SDC (OSDC)	The application of SDC methods to potential publications after the analysis has been carried out, to guard against the residual risk
Principles-based	A regulatory regime or operating model where 'principles' (what you are trying to achieve) are the basis for planning. Rules are designed to implement principles but can be changed if inconsistent. A principles-based system does not specify how a goal is to be achieved, only what the goal is. For example, a data protection regime could specify that the confidentiality of the individual is protected, but without specifying whether that occurs through anonymization or other methods.
Public use file (PUF)	Data file without restrictions on use or onward access
Raw data	The source data
Remote job server (RJS)	A system allowing a range of complex analyses to be carried out, not just tabulations, without seeing the source data; a table server is a remote job server that has only one function
Research data centre (RDC)	A restricted access facility where users can manipulate the source data without restriction as if on their own computers; but the environment is made secure so that users cannot bring information into or take data out of the facility without approval, and additional services (such as internet access) are normally very restricted, typically provided by on-site access, where the facility is hosted on the organisation's premises
Rules-based	A regulatory regime or operating model where explicit rules are the basis for planning. The rules may specify how individuals and organisations should act, or (for example, defining 'anonymisation' and specifying what can be done with data that has or has not been anonymized).
Scientific use file (SUF)	The data file which retains some non-negligible confidentiality risk and so, therefore, has circulation restricted to authorised users for specific research purposes
Secure use file (SecUF)	The data file which contains non-negligible confidential information therefore circulation and use is restricted to authorised users in controlled facilities
Sensitive data	Data where release to an unauthorised person is likely to cause nonnegligible harm or distress to the data subject; for this report, we assume that all sensitive data is also confidential

Statistical disclosure control (SDC)	Applying statistical measures to (a) determine if there is a substantive risk of unauthorised disclosure in a dataset or publication, and (b) make changes to the data or publication to reduce that risk
Synthetic data	Generated data that can replace or augment sensitive source data
Table server	A system that allows users to generate their tables from the data flexibly, but without seeing the source data; a form of distributed access
Unauthorised disclosure	The unauthorised release of information about an identified data subject
Virtual RDC or Remote RDC (vRDC)	An RDC where technology is used to provide equivalent security to a physical site and to separate the RDC from the actual location of the data