PATHFINDER STUDIES' OUTPUTS WORKSHOP

SUMMARY REPORT

2023-2024

This report was written by Health Data Research Global, Health Data Research UK in collaboration with its partners, The Global Health Network, Fiocruz and icddr,b.

Table of Contents

1. Background

The 'New Ecosystem' partnership programme was established to build and strengthen the global health research and data science ecosystem, funded by the Bill & Melinda Gates Foundation. This collaborative partnership involves The Global Health Network, Fiocruz, Icddr,b, Africa CDC and HDR Global (the latter convened by Health Data Research UK). A key component of the partnership programme is to use exemplar or ***host projects*** that are using data science approaches to address major health research questions and set up associated ***pathfinder studies*** to help identify challenges and solutions that help accelerate the health data research cycle and enable the host project to provide valuable new insights. These pathfinder studies seek to share their findings and solutions within the domain of open science research, empowering other research teams to access and utilise the valuable tools and methodologies generated, as well as to build data science and health research capacity.

Through this partnership programme, two pilot host projects and associated pathfinder studies have been established:

- Linking datasets to understand regional differences in micronutrient deficiency in children-to inform policy and practice, Bangladesh (icddr,b).

- Effectiveness of COVID-19 Vaccination in Brazil Using Mobile Data (DP-EFFECT-BRAZIL), Brazil, Fiocruz.

Several key resources have been already developed for and from the pathfinder studies, including concept notes, a tracker tool and a policy brief. Following review by the host project and pathfinder study teams, it was collectively recognised that the next step was to reach agreement on expected pathfinder study outputs, as well as the benefits they bring to the host projects, their institutions, and a wider range of stakeholders. Therefore, the partners worked together to design an online workshop, consisting of two sessions on two consecutive days with the aim to:

- Gather consensus on the outputs of the pathfinder studies.
- Promote a shared understanding of the pathfinder studies' components of the New Ecosystem programme with respect to outputs/deliverables, including tools, methodologies and other approaches.

- Promote a shared understanding of the New Ecosystem programme implementation and how the pathfinder studies are contributing to the delivery of the other outcome/impact areas.
- Identify the challenges and solutions already identified by the pilot host projects and associated pathfinder studies.
- Identify the beneficiaries of the pathfinder studies and evaluate the broader advantages/benefits they bring along with next steps.

2. Scope of this summary report

This report aims to provide a comprehensive overview of the prominent themes that emerged during the two workshop sessions, which focused on four critical areas. Each of the areas covered sought to identify the core outputs that can be generated from pathfinder studies, the challenges encountered and potential solutions, all of which have been captured.

In addition, the report captures the benefits of carrying out pathfinder studies, as well as the benefits that key stakeholders - including communities - stand to gain from both host projects and pathfinder studies.

The following four areas covered during the workshop sessions were:

- Accelerating research insights in priority health areas
- Capacity building in data science
- Accelerating access, sharing and re-use of data
- Stakeholder and community engagement

3. Pathfinder Studies Outputs, Challenges and Solutions

Overall, the workshop sessions identified that pathfinder studies can play a significant role in advancing health research and enhancing staff development across multiple levels, encompassing individuals, research teams and benefiting institutions. The following paragraphs capture key findings that arose from the brainstorming activity across the different themes / areas. In addition, **Figure 1** in the appendix illustrates a schematic diagram of the key findings that arose from the brainstorming activity.

3.1 Accelerating Research Insights in Priority Health Areas: Outputs, Challenges and Solutions

The main **outputs** from pathfinder studies identified as accelerating the generation of research insights in priority health research areas included:

- The development of a robust research plan, processes, methodologies, and monitoring/tracking tools for all stages of the health data cycle within the host project

- Plans for analysis, workflows, process maps/diagrams, along with standardised protocols for the host study.

- Development of robust data collection, harmonisation and linkage methodologies and practices

- Pathfinder studies help accelerate the health data research cycle and therefore the generation of new evidence and insights to inform health policy.

- Development of focused training/workshops or Institutional curriculums to upskill in health data science.

- Establishment of effective communication and monitoring between and within teams.

The main **challenges** that were identified for pathfinder studies in accelerating research insights in priority health areas include:

- Challenges of retrospective studies in tracking and recording accurate data capture, processes, and methodologies.

- Data sharing issues.

- Communication issues between host and pathfinder teams.

- Lack of funding.

- Institutional challenges due to lack of support, management skills and work overload.

- Challenges in tracking processes and impact.

- Challenges with study timelines.

The main potential **solutions** that were identified in accelerating research insights in priority health areas include:

- Creating new funding opportunities for exploring methodology in data analysis and guidelines.

- Efficient documentation of methodologies and processes, along with tracking of research progress.

- Development of toolkits, processes, methodologies, and templates to track and communicate outputs and impact.

- Outline clear research benefits to the Institution.

- Effective communication within the teams and between stakeholders.

- Funding and Institutional support in focused training.

- Creating a community culture and engaging stakeholders in effective management and communication practices.

3.2 Data Science Capacity Building: Outputs, Challenges and Solutions

The main **outputs** identified for data science capacity building included:

- Workshops in R programming or other relevant tools.

- Research-focused training via workshops/training or data clubs and clinics.

- Institutional courses.

- Study focused workshops in all the different steps of data science cycle.

- Mentorship and tutoring programmes.

- Data knowledge exchange/sharing sessions with other groups.

- Data science learning hubs.

- Training on publishing data driven papers.

- Resource management and utilisation.

The main **challenges** identified for data science capacity building include:

- Identifying and addressing different skills requirements for different role holders in pathfinder study and host project teams

- Lack of human resources and dedicated teams to provide data science capacity strengthening.

- Challenges in finding resources for capacity building that are appropriate for the context and research studies.

- Lack of Institutional support and funding.

- Lack of communication skills to be able to deliver workshops/training properly.

- Lack of training opportunities.

The main **solutions** that were identified in data science capacity building include:

- Funding for data science capacity building/training focused on research needs.

- Funding a dedicated team to provide training.

- Documentation of barriers or challenges faced can help identify solutions.

- Networking / knowledge hubs to enable knowledge/expertise sharing and lessons learned from similar studies.

- Pathfinder study teams can help organise capacity building activities internally or showcase what was produced by the host study and would be useful externally.

3.3   Accelerating Data Access, Sharing and Re-use: Outputs, Challenges and Solutions

The main **outputs** that were identified for accelerating data access, sharing and re-use included:

- Data sharing agreements, institutional roadmap for data re-use and monitoring.

- Data standardisation practices across the different stages of the health data cycle.

- Data Management Plan/templates and tools.

- Access to technological tools to accelerate research and create data catalogues and repositories.

- Communication and knowledge sharing of data-driven studies between global research communities.

- Focused funding on data sharing initiatives.

The main **challenges** that were identified in accelerating data access, sharing and re-use*: included:

- Challenges with data linkage and data sharing due to PIs/ Institutes and research teams not keen to share data due to concerns with data use/handling.
- Ethical, legal and governance issues around data access, sharing and re-use.

- Challenges with finding or accessing of datasets.

- Lack of standardised and validated processes for each stage of the health data research cycle and for different types of data to enable quality assurance.

- Challenges around expertise in the use of data related tools/technologies.

- Lack of funding for relevant initiatives

- Attributions and recognition.

The main **solutions** that were identified in accelerating data access, sharing and re-use*: included:

- Development of frameworks, including guidelines and standards to enable trustworthy data access and sharing with input by legal team and advisory groups.

- Promote inclusivity and communication between teams.

- Data Standardisation/Harmonisation

- Community Engagement on data use for research purposes

- Promote new approaches and technical understanding of data sharing.

- Knowledge exchange to create a culture of data sharing and open science principles.

- Funding opportunities

- Outline clear benefits on data sharing for national and regional governments that are involved and benefitting from relevant research.

3.4 Stakeholder Engagement: Outputs, Challenges and Solutions

The main **outputs** that were identified in stakeholder engagement, including local communities, included:

- Work with local communities to identify health research priorities and keep the communities engaged throughout the research process.

- Build trust with the community/participants and work towards combating fake news.

- Initiatives on data science capacity building in the community.

- Informing future research through policy briefings.

- Publications on lessons learned and the methodologies used, and approaches taken.

- Communication on research innovation and research Impact.

The main **challenges** that were identified in stakeholder – particularly community engagement - included:

- Lack of understanding amongst residents in local communities who are participating in the studies and their fear of serving as a scientific test.

- Challenges for research teams and wider stakeholders in understanding the Pathfinder study concept.

- Challenges with summarising and communicating research findings to the community/stakeholders.

- Challenges in keeping communities interested and engaged.

- Data sharing challenges due to privacy concerns.

- Challenges in staff retention.

The main **solutions** that were identified for stakeholder and community engagement included:

- Transparency on benefits, risks and expectations across all parties involved.

- Actively engage with the community and understand their needs.

- Clear understanding of the needs of teams involved in research.

- Effective research communications across stakeholders and beyond.

- Campaigns to combat fake health research news in the community.

- Efficient strategy development across key areas.

4. Benefits of Pathfinder Studies Outputs

- They create a sustainable data science infrastructure around host research projects, as well as for their institutes and enable acceleration of data science approaches for host studies to deliver critical new health insights.

- Act as exemplar studies and reduce the risks and challenges of similar future studies.

- Share innovation and knowledge/expertise with the wider research community on data science practices.

- Enable wider research progress in addressing specific health challenges by facilitating research studies in other regions ethnic/groups or communities.

- Provide insights that Improve healthcare for the community.

- Build capacity and provides focused data science training programmes at an individual and institutional level to benefit all individuals involved in relevant research.

- Strengthen and clarify legal frameworks on data access and sharing to enable a culture of data sharing and adopting open science principles.

- Influence policy making in relation to two areas 1) Accelerates the heath data research cycle to produce new insights, to inform health policy; 2) Provide policymakers with information on the challenges and solutions to enable effective data infrastructure, access and sharing on a national and multinational level.

- Publications produced that communicate research findings to the wider scientific community.

5. Beneficiaries of Pathfinder Studies Outputs

- Host project and pathfinder study team members of research institutes and organisations involved in relevant research.
- Research and ethic board and study committees.
- The wider scientific community.
- Healthcare systems and members of staff within these.
- Policymakers and governments.
- Patients affected by research hypothesis.
- Wider public and vulnerable communities.

- Data owners, producers and custodians
- Funding bodies/Donors.
- NGOs involved with relevant research.

6. Conclusion

The workshop sessions highlighted that the pathfinder studies yield a diverse array of outputs and benefits, particularly for individual researchers, host project/pathfinder study teams and their institutions. Collectively, these contribute to the evolution and strengthening of a health data science ecosystem and the host projects and pathfinder studies can serve as exemplars for future studies employing similar methodologies. Furthermore, the inherent potential to establish legal, ethical and governance frameworks capable of overcoming data sharing challenges is a promising prospect that can accelerate priority health research and bring new evidence to improve health outcomes for everyone, everywhere.

The main challenges encountered, aside from legal constraints and data sharing limitations, revolve around the ambiguity associated with the pathfinder study concept and the absence of dedicated funding to support its execution. This issue arises from the challenge faced by institutions and funders in recognising the tangible benefits of putting in place effective data infrastructure, governance, and data sharing approaches, which hinders the effective delivery of both host projects and pathfinder studies.
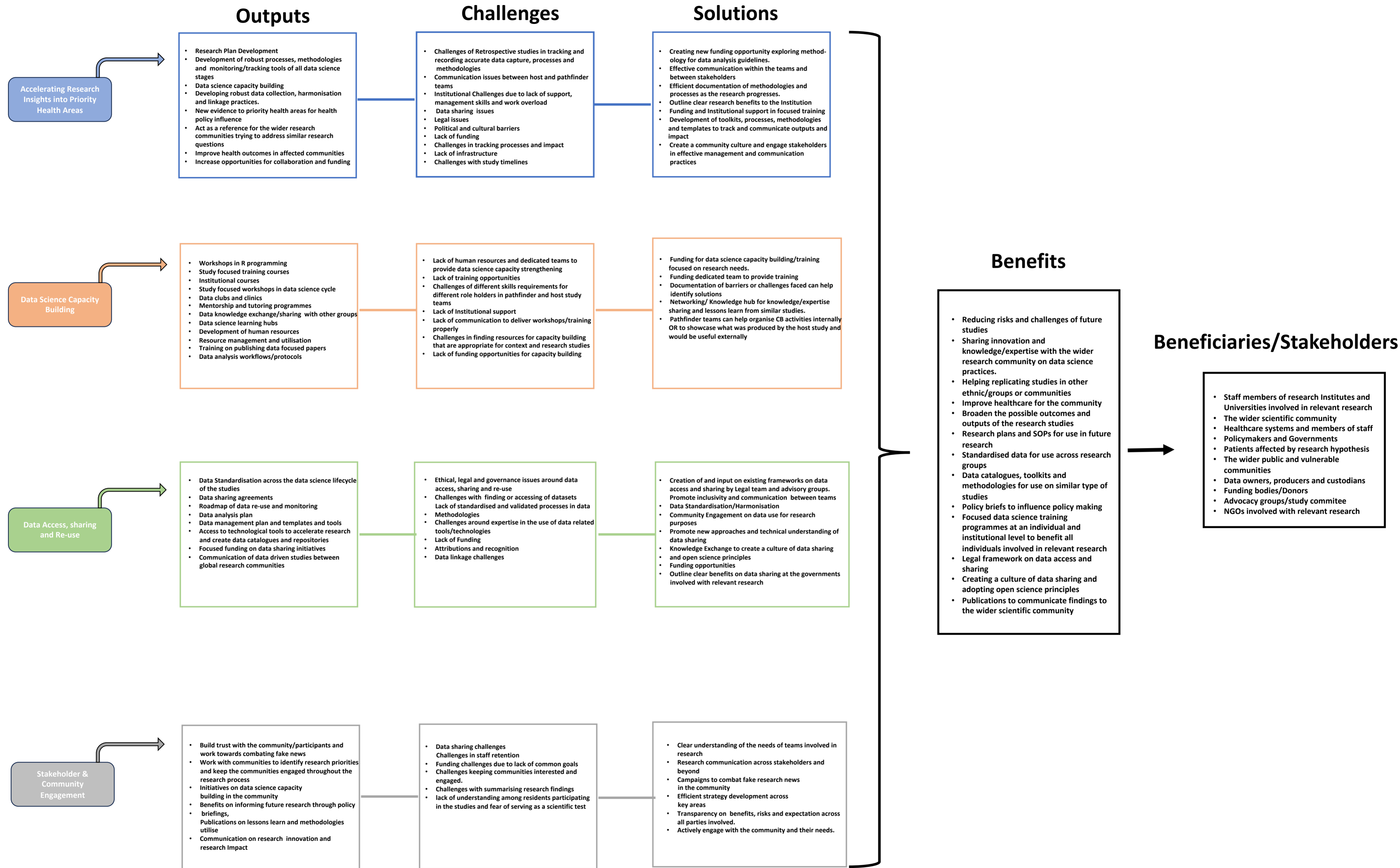
7. Appendix

Outputs

Accelerating Research Insights into Priority Health Areas
- Research Plan Development
- Development of robust processes, methodologies and monitoring/tracking tools of all data science stages
- Data science capacity building
- Developing robust data collection, harmonisation and linkage practices.
- New evidence to priority health areas for health policy influence
- Act as a reference for the wider research communities trying to address similar research questions
- Improve health outcomes in affected communities
- Increase opportunities for collaboration and funding

Data Science Capacity Building
- Workshops in R programming
- Study focused training courses
- Institutional courses
- Study focused workshops in data science cycle
- Data clubs and clinics
- Mentorship and tutoring programmes
- Data knowledge exchange/sharing with other groups
- Data science learning hubs
- Development of human resources
- Resource management and utilisation
- Training on publishing data focused papers
- Data analysis workflows/protocols

Data Access, sharing and Re-use
- Data Standardisation across the data science lifecycle of the studies
- Data sharing agreements
- Roadmap of data re-use and monitoring
- Data analysis plan
- Data management plan and templates and tools
- Access to technological tools to accelerate research and create data catalogues and repositories
- Focused funding on data sharing initiatives
- Communication of data driven studies between global research communities

Stakeholder & Community Engagement
- Build trust with the community/participants and work towards combating fake news
- Work with communities to identify research priorities and keep the communities engaged throughout the research process
- Initiatives on data science capacity building in the community
- Benefits on informing future research through policy briefings,
  Publications on lessons learn and methodologies utilise
- Communication on research innovation and research Impact

Challenges

Accelerating Research Insights into Priority Health Areas
- Challenges of Retrospective studies in tracking and recording accurate data capture, processes and methodologies
- Communication issues between host and pathfinder teams
- Institutional Challenges due to lack of support, management skills and work overload
- Data sharing issues
- Legal issues
- Political and cultural barriers
- Lack of funding
- Challenges in tracking processes and impact
- Lack of infrastructure
- Challenges with study timelines

Data Science Capacity Building
- Lack of human resources and dedicated teams to provide data science capacity strengthening
- Lack of training opportunities
- Challenges of different skills requirements for different role holders in pathfinder and host study teams
- Lack of Institutional support
- Lack of communication to deliver workshops/training properly
- Challenges in finding resources for capacity building that are appropriate for context and research studies
- Lack of funding opportunities for capacity building

Data Access, sharing and Re-use
- Ethical, legal and governance issues around data access, sharing and re-use
- Challenges with finding or accessing of datasets
- Lack of standardised and validated processes in data
- Methodologies
- Challenges around expertise in the use of data related tools/technologies
- Lack of Funding
- Attributions and recognition
- Data linkage challenges

Stakeholder & Community Engagement
- Data sharing challenges
  Challenges in staff retention
- Funding challenges due to lack of common goals
- Challenges keeping communities interested and engaged.
- Challenges with summarising research findings
- lack of understanding among residents participating in the studies and fear of serving as a scientific test

Solutions

Accelerating Research Insights into Priority Health Areas
- Creating new funding opportunity exploring methodology for data analysis guidelines.
- Effective communication within the teams and between stakeholders
- Efficient documentation of methodologies and processes as the research progresses.
- Outline clear research benefits to the Institution
- Funding and Institutional support in focused training
- Development of toolkits, processes, methodologies and templates to track and communicate outputs and impact
- Create a community culture and engage stakeholders in effective management and communication practices

Data Science Capacity Building
- Funding for data science capacity building/training focused on research needs.
- Funding dedicated team to provide training
- Documentation of barriers or challenges faced can help identify solutions
- Networking/ Knowledge hub for knowledge/expertise sharing and lessons learn from similar studies.
- Pathfinder teams can help organise CB activities internally OR to showcase what was produced by the host study and would be useful externally

Data Access, sharing and Re-use
- Creation of and input on existing frameworks on data access and sharing by Legal team and advisory groups. Promote inclusivity and communication between teams
- Data Standardisation/Harmonisation
- Community Engagement on data use for research purposes
- Promote new approaches and technical understanding of data sharing
- Knowledge Exchange to create a culture of data sharing and open science principles
- Funding opportunities
- Outline clear benefits on data sharing at the governments involved with relevant research

Stakeholder & Community Engagement
- Clear understanding of the needs of teams involved in research
- Research communication across stakeholders and beyond
- Campaigns to combat fake research news in the community
- Efficient strategy development across key areas
- Transparency on benefits, risks and expectation across all parties involved.
- Actively engage with the community and their needs.

Benefits
- Reducing risks and challenges of future studies
- Sharing innovation and knowledge/expertise with the wider research community on data science practices.
- Helping replicating studies in other ethnic/groups or communities
- Improve healthcare for the community
- Broaden the possible outcomes and outputs of the research studies
- Research plans and SOPs for use in future research
- Standardised data for use across research groups
- Data catalogues, toolkits and methodologies for use on similar type of studies
- Policy briefs to influence policy making
- Focused data science training programmes at an individual and institutional level to benefit all individuals involved in relevant research
- Legal framework on data access and sharing
- Creating a culture of data sharing and adopting open science principles
- Publications to communicate findings to the wider scientific community

Beneficiaries/Stakeholders
- Staff members of research Institutes and Universities involved in relevant research
- The wider scientific community
- Healthcare systems and members of staff
- Policymakers and Governments
- Patients affected by research hypothesis
- The wider public and vulnerable communities
- Data owners, producers and custodians
- Funding bodies/Donors
- Advocacy groups/study commitee
- NGOs involved with relevant research

Figure 1: Schematic diagram of the key findings that arose from the brainstorming activity