Scoping Report:
Data Science Challenge Prizes for Health
Wellcome Data For Science and Health

Programme Scoping

Executive summary 2
Background Information 3
Prize Programme Focus 5
Key Decisions 6
Design principles 8
Glossary of terms 9
Challenge Topic Summaries 10
Delivery Recommendations & Timeline 18
Governance, Reporting & Evaluation 21
Programme Budget 26
Communications Strategy Summary 28

Challenge Prize Specific Scoping

Mental health data challenge: Detailed Scoping 30
Snakebite health data challenge: Detailed Scoping 59
Drug resistance infections (DRI) health data challenge: Detailed Scoping 84
Urban health data challenge: Detailed Scoping 117

Appendices 152
Executive summary

This document is the result of a 9 month scoping exercise carried out from September 2019 to May 2020 by Open Data Institute on behalf of the Data for Science and Health team at Wellcome. The scoping engaged with stakeholders from Wellcome and across the healthcare domain to ensure the challenges prizes are structured to align with and deliver impact against Wellcome's priorities. The method followed can be found in the Data Challenge Prizes for Health Playbook\(^1\) which has been developed in parallel to the scoping activities.

The document includes research on four potential challenge prize topics which have been evaluated across different regions. These were prioritised from a longlist of potential topics\(^2\) identified through engagement with a broad group of health and data specialists, using the criteria around the relevance and appeal of the challenge topic, how sustainable the impacts might be, the relevance of data science and data sharing, the human impact, how measurable the impact could be and potential ethical, cultural and legal issues\(^3\).

The four shortlisted challenge topics which have been researched in detail are:

<table>
<thead>
<tr>
<th>Topic</th>
<th>Region under consideration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data science for more effective mental health treatments for young people</td>
<td>UK</td>
</tr>
<tr>
<td>Data science for reducing snakebite mortality</td>
<td>India</td>
</tr>
<tr>
<td>Data science for improving antibiotic prescriptions</td>
<td>Kenya, Malawi, Uganda</td>
</tr>
<tr>
<td>Data science for public health risks in urban environment</td>
<td>Nigeria or Kenya</td>
</tr>
</tbody>
</table>

This document contains a summary of these topics, along with information on how the programme of challenge prizes as a whole could run with regard to governance, costings, sequencing and risks.

Supporting this programme overview doc, each of the specific challenge topics has a detailed scoping document containing information on:

- The desired impact of the Challenge topic and how it will be achieved and measured
- The current state of the data ecosystem surrounding the Challenge topic
- The political, legal and regulatory environments in the target areas for the Challenges
- The stakeholders involved in the Challenge topic
- Challenge plans, including their reward structure
- The risks and opportunities of running a Challenge Prize around the topic

---

\(^1\) Appendix 6  
\(^2\) Appendix 7  
\(^3\) The full scoring sheet for the longlist can be found at Appendix 8
The content of this Challenge programme scoping report aligns closely with the strategic objective and direction of Wellcome's current priority areas. However some of the recommendations may be impacted by the consequences of COVID-19 in ways which cannot be reliably predicted at this time.

Background Information

Wellcome's latest priority area, Data for Science and Health aims to equip and motivate data scientists to innovate with health data for public good, while engaging society to build people's trust, understanding and participation in health data innovation. As part of this larger programme of work, they want to build multidisciplinary collaborations through a global programme of Data Science Challenge Prizes.

A Challenge Prize is one way of helping to drive innovative uses of data. Challenge prizes resolve some of the barriers to innovation by offering an incentive to a broad range of people to produce innovative solutions.

A Challenge Prize should start with a well-defined problem. The types of problems that are most suitable for a prize centred around health data innovation have the following characteristics:

1. **Problems must have direct societal relevance**
   A good challenge problem needs to attract a wide pool of potential innovators. It therefore needs to have wide societal relevance. This will mean that different practitioners from different disciplines will understand the problem and be motivated to try and contribute to solving it.

   Health data challenge prize problems need to be big enough to attract participants who can think about the project differently and bring different approaches and techniques to the solutions. However, make sure that it’s not so big you won’t be able to clearly describe and measure the impact of your challenge prize.

2. **Problems which can be addressed through the use of data**
   A good health data challenge prize problem will focus on addressing an existing health problem through the use of data in a new way. This could mean using new techniques in processing and using health data, or the combination of health data with other less traditional sources.

   The topic area and region you are running it in therefore needs to have at least some data available specific to the problem area and a technical and regulatory data infrastructure to enable access to that data by challenge participants. Health data is highly sensitive so access is often restricted and subject to rigorous governance. You will need to be able to work with data stewards to enable access or the creation of synthetic data to support participants.
3. **Problems need to have measurable solutions**  
The health challenge prize problem needs to have a solution that is measurable and that enables you to set time-bound targets to measure the progress you’re making. In order to demonstrate the value of these types of activities, you need to have something to show at the end of the prize which interests other funders and partners to invest and galvanise further work on the projects.

4. **Problems need to have coordinated efforts around them**  
Some problems will have lots of people working away at them from different angles and perspectives. This is great in some ways as it means there’s clearly a problem to be solved, but it might also mean that efforts aren’t as coordinated or progress as quickly as they could if people worked together. Running a challenge prize allows you to invite lots of different people to apply and provides an opportunity to bring different skill sets together.

These Prizes are intended to address the following issues that have been identified by Wellcome:

1. There are too few data scientists working with health data for public good.
2. Trust in the use of data for health innovation is fragile.
3. The health/data community lacks guidance around running health data innovation projects, including guidance around effective methods for designing health data innovation projects.

Wellcome intends the Data Science Challenge Prize programme to enable the use of data science to address significant health problems, with the aim of encouraging collaborations across a range of partners and multidisciplinary teams and increasing the capability of people working with health data.

Each Challenge will take place in a different region, with at least one Challenge Prize in the UK, one in sub-Saharan Africa and one in India. The Challenges themselves will target different points within the system where greater access to information can make an impact, such as data analysis to help inform policy makers prioritise interventions; applications that support healthcare professionals; or tools to support health networks.

The Challenges will include an element of competition, providing an opportunity to raise the visibility of data science and generate excitement. Most importantly, each Challenge Prize will deliver a health impact. This will allow the programme to inspire the public, clinicians and decision-makers with opportunities for responsible health data innovation, with each Challenge acting as high-profile exemplar. It will also motivate data scientists to work with health data, by demonstrating that the health sector has exciting technical challenges to tackle that can also lead to social benefit.

The first Challenge Prize could be initiated on the basis of the insights in this scoping report. The recommendation is that Wellcome should briefly revisit the Challenge Prize proposals in the second year of the programme to review and select the final Challenge Prizes. This will allow lessons learned from the first Challenge Prizes to be responded to and any changing circumstances can be taken into account.
Prize Programme Focus

What is the problem you are trying to address for the health data challenge series programme as a whole?

There is not enough health data innovation happening in the health sector which is ethical, provides balanced equity to all, and engages with the communities affected and influenced by it.

Sub-problems:
- There are too few data scientists working with health data for public good.
- Trust in the use of data for health innovation is fragile.
- The health/data community lacks guidance around running health data innovation projects, including guidance around effective methods for designing health data innovation projects.

Challenge Prize Programme Vision
This programme will demonstrate how to create data innovations for health that are ethical, provide balanced equity to all, and engage with the communities affected and influenced by it. In doing so, it will increase data science capacity, public trust and guidance.

Theory of Change
A full theory of change for the programme can be found below. It shows the relationship between the types of activities the challenge prizes will undertake, the types of outputs that are likely to result from these activities, the outcomes these outputs will have, and the overall impact that will be achieved.

In the programme level theory of change there are two streams of activities. The first is the running of the challenge prizes. The corresponding impacts for these activities are stated to be more data scientists motivated and able to work in healthcare, increased public trust in the use of data for health innovation, and increased capacity within the health sector to use data by building networks with organisations who can provide that capability.

The second stream is to develop resources for designing and running health data innovation projects. The resulting impact for this stream aims to be increased expertise in the health/data community to run health data innovation projects.

To support this second stream, ODI have compiled a playbook providing a clear methodology for running data science challenge prizes in health to support other organisations interested in the approach. This can be found in Appendix 6 and will continue to be iterated on by ODI.
<table>
<thead>
<tr>
<th>Activities</th>
<th>Outputs - what is produced through the activities</th>
<th>Outcomes - what do the outputs achieve (NB: used to inform judging criteria for prizes)</th>
<th>Impact - long term, outcomes contribute to a wider set of interventions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Run challenge prizes which engage: Data Scientists, Tech community, Health data community, Research community, Patients, General public (citizens)</td>
<td>Build multidisciplinary partnerships with universities, NGOs, other relevant institutions in designing and running the challenges</td>
<td>Tools and applications are developed which improve the capability of the field or region in response to the stated challenge in one or more of the following ways: 1. Individual care 2. Understanding disease 3. Evaluating policy 4. Planning health services 5. Diagnosis 6. Treatment and prevention 7. Patient safety</td>
<td>Increased awareness in the data science community of the value they can add in healthcare through demonstrating that: 1) There are opportunities for data science to find meaningful and impactful solutions to healthcare problems in innovative ways 2) Health tech innovation built in an ethical, equitable manner which prioritises engaging with communities affected by the technology throughout development is more effective. 3) There are opportunities for collaboration, integration and recognition for data science in health</td>
</tr>
<tr>
<td></td>
<td>Provide guidance and support to all participants to access medical data in a timely manner</td>
<td></td>
<td>1) Health data innovation can be impactful</td>
</tr>
<tr>
<td></td>
<td>Identify and enable data access opportunities for specific challenges</td>
<td></td>
<td>2) Health data innovation can be done ethically and responsibly</td>
</tr>
<tr>
<td></td>
<td>Promote the challenge and engage with challenge participants</td>
<td></td>
<td>3) Health data innovation can be done in an inclusive way</td>
</tr>
<tr>
<td></td>
<td>Provide seed funding</td>
<td></td>
<td>4) Health data innovation can be done in a way where it benefits everyone equally (healthcare providers, patients, citizens etc...)</td>
</tr>
<tr>
<td></td>
<td>Judge entries and give prize</td>
<td>Networks/ relationships are established, supported and reinforced: across tech, academia and healthcare</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Provide mentoring</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Provide sustainability support for entrants</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Strengthen the community of data scientists working on the challenges and beyond through peer network support</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Communicate/ share stories with the wider public about the challenges</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Engage with ’end user’ groups (patients/ clinicians/ healthcare providers/ researchers/ citizens) in order to inform/ co-produce the prize entries</td>
<td>Public-facing stories are published about the use of health data for innovation</td>
<td></td>
</tr>
<tr>
<td>Develop resources for designing and running health data innovation projects</td>
<td>Test methods for identifying open innovation models for specific challenge topics</td>
<td>Best practice guidance is developed and shared about how to engage with the public when building these kinds of systems</td>
<td>The broader health/data community has access to reusable methods and approaches for: 1) Identifying open innovation models for health data</td>
</tr>
<tr>
<td></td>
<td>Design playbook for data challenge topic readiness evaluation</td>
<td></td>
<td>2) Engaging citizens and patients in health data innovation projects</td>
</tr>
<tr>
<td></td>
<td>Test different models for participatory engagement throughout the challenge prizes</td>
<td></td>
<td>3) Using data in health data innovation projects</td>
</tr>
<tr>
<td></td>
<td>Develop case studies for participatory engagement methods</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Test different technology and governance approaches for health data innovation through the prize entries</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Develop demonstrators of different technology and governance approaches</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1 Categories taken from Understanding Patient Data’s ‘ways data can help’: [https://understandingpatientdata.org.uk/why](https://understandingpatientdata.org.uk/why)
## Key Decisions

This document is an agnostic reflection of the scoping research. As such there are several key decisions to be made based on this information. Below is an overview of these decisions and a recommendation from ODI.

<table>
<thead>
<tr>
<th>Decisions</th>
<th>Recommendation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Does Wellcome wish to run challenge prizes on all four topic areas?</td>
<td>Proceed with the first two challenges (Mental Health and Snakebite) and build in review points in which to re-evaluate the other topics with a view to commissioning them accordingly. This also provides the opportunity to pivot and re-run a challenge topic for a second round or in another location and allow for the impacts of COVID-19 to be taken into consideration.</td>
</tr>
<tr>
<td>Will the other Wellcome Priority Area teams co-fund these challenges?</td>
<td>Each of the challenge topic areas in this document relate to a Wellcome Priority Area and have been designed with those strategic priorities in mind. Mental Health, Snakebite and Drug Resistant Infections have expressed interest in co-funding the respective challenge prizes. Urban Health is subject to the strategic prioritisation currently occurring but is in alignment with the existing Our Planet Our Health Priority Area.</td>
</tr>
<tr>
<td>What location is best suited to run the Drug Resistant Infections Challenge Prize?</td>
<td>The scoping for Drug Resistant infections has information on three countries; Malawi, Uganda and Kenya. Ghana has also been suggested as a suitable country to run such a prize in. As the proposal is reliant on data which is to be collected as upcoming work across the DRI and Public Engagement teams, it is recommended that choosing the most appropriate location should be done following a review of that data and a period of engagement with the Ministries of Health in the various countries.</td>
</tr>
<tr>
<td>What location is best suited to run the Urban Health Challenge Prize?</td>
<td>The scoping for the Urban Health topic includes information on Kenya and Nigeria. The key decision for this topic is where the delivery team can get the most engagement with national and state governments as they are the ideal end users/ beneficiaries of the outputs and therefore need to be open to adopting them. Through the course of the scoping the contacts in Nigeria have been highly forthcoming, but Wellcome has many established contacts in Kenya which could be used to create these connections.</td>
</tr>
<tr>
<td>Has the scoping focused on the right questions within the subject area?</td>
<td>The scoping has covered a broad range of topics with as much specificity as possible within the allocated time. Once Wellcome has taken the decision to commit to one or multiple challenge topics, the recommendation is to engage in a set-up phase which allows the team to verify the specific challenge question with key stakeholders. This will also enable the impacts of COVID-19 to be taken into consideration.</td>
</tr>
</tbody>
</table>
Design principles

The challenges prizes have been scoped with the following principles in mind:

**Address a significant health problem**
Focus on health problems which can be addressed using data science but that also deliver tangible, visible impact to the wider health sector and are understandable to a wide range of potential participants and stakeholders.

**Use existing data, do not focus on making new data**
Where possible demonstrate that there are already a huge number of opportunities for the use of data science in the health sector with existing data. This will focus the Challenges Prizes on innovation rather than creating data infrastructure.

**Co-create with the community**
Transparency is an important part of building public trust, but trust also requires participation. Therefore the patient groups and communities most affected by the challenge also need to be part of co-designing and delivering the prize as expert stakeholders and judges.

**Enable all participants**
One criticism of challenge prizes can be that they require a large amount of free labour for the possibility of financial remuneration. In many cases with early stage health companies who are most agile and innovative are pre-revenue and rely heavily on grants, subsidies and investment. Similarly academic, civil society and public participants often struggle to engage with prizes due to a lack of funding. Therefore in order to enable everyone who has relevant experience and innovative ideas to take part, it is important to offer seed funding for all stages of participation.

**Build capacity**
Where possible, aim for the challenge prizes to build data science capacity across a peer network of experts by bringing communities together and providing training and mentorship through the prize. This is another way of ensuring sustainability and meeting the objective of increasing data science capability across the health sector.

**Make it generalisable**
The Challenge Prizes should be relevant to a range of places meaning that innovations created can be replicated or scaled elsewhere, or the Challenge Prize itself can be repeated in another location.

**Be driven by the end user**
In order to create things which actually answer the challenge topic in a practical and usable way, it is crucial to position the challenges around the needs of those who are directly affected by the issue.

**Make the impact visible**
Show the impact of these innovations through measurable change and human stories. Show the impact of both the innovations produced in the challenge prizes and the wider impacts of the programme of challenge prizes.

**Create sustainable solutions**
Sustainability is crucial to ensuring impact from the innovations generated through the Challenge Prizes. Consider how they will be adopted and sustained after the term of the prize.
Glossary of terms

**Challenge prize**
An open prize in which a financial reward is levied in order to motivate a wide pool of innovators to respond to a specific social or technical ‘challenge’ or problem which isn’t being solved with usual innovation methods and therefore would benefit from a wide pool of innovators looking to tackle the issue with unorthodox thinking or methods.

**Challenge programme**
The series of phases and activities required to execute a challenge prize. This usually includes some combination or variation of scoping, design and development, pre-launch, launch, applications, team and idea development, incubation, assessment and judging, award, and post-award phases.

**Challenge topic**
An identified problem or problem area for which a challenge prize is best suited to address. Challenge topics must have direct societal relevance, are addressable through the use of data and need to have measurable solutions.

**Data access initiative**
A joint activity and programme of work that aims to address a specific challenge through increasing access to data

**Data innovation**
A new idea, method, or device that relies upon access to data to deliver value

**Data ecosystem**
A data ecosystem consists of data infrastructure, and the people, communities and organisations that benefit from the value created by it

**Data ecosystem map**
A data ecosystem map illustrates the different actors in a data ecosystem, and how value is exchanged across it.

**Data ethics**
Data ethics is a branch of ethics that evaluates data practices with the potential to adversely impact on people and society – in data collection, sharing and use

**Data infrastructure**
Data infrastructure is made up of data assets, standards, technologies, policies and the organisations that steward and contribute to them.

**Data inventory**
A list of datasets with metadata that describes their contents, source, licensing and other useful information.
Data landscaping
A data landscape is a process that can help discover what data is out there, how much could it be used, accessed and shared, to solve a given challenge (or in a particular context), as well as identifying any issues and enablers in doing so. It helps identify key data publishers and users as well.

Data science challenge prize
A challenge prize in which the innovation function is driven by data science. This usually includes some combination of access, collection, analysis, use and sharing of data in order to deliver value.

Data steward
Person or organisation who collects, maintains and shares data

Data user
Person or organisation who purposefully uses or transforms data towards a specific function

Open data
Open data is data that anyone can access, use and share.

Open innovation
An innovation process based on purposefully managed data, information and knowledge flows across organisational boundaries in order to gain knowledge and ideas from external sources.

Personal data
Personal data is information that relates to an identified or identifiable individual.

Theory of change
A description, illustration and rationale of how and why a desired change can happen in a particular context

Challenge Topic Summaries
This section includes a short summary of the four topic areas which have been scoped. Full detailed information is available later in the document.

<table>
<thead>
<tr>
<th>Topic</th>
<th>Region under consideration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data science for more effective mental health treatments for young people</td>
<td>UK</td>
</tr>
<tr>
<td>Data science for reducing snakebite mortality</td>
<td>India</td>
</tr>
<tr>
<td>Data science for improving antibiotic prescriptions</td>
<td>Kenya, Malawi, Uganda</td>
</tr>
<tr>
<td>Data science for public health risks in urban environment</td>
<td>Nigeria or Kenya</td>
</tr>
</tbody>
</table>
Data Science for more effective mental health treatment for young people: Summary

**How can we support the development of more effective interventions for addressing anxiety and depression in young people and adolescents?**

Depression and anxiety, two of the most commonly diagnosed mental health disorders, are often seen as a catch-all definition for someone’s experience. But just as people are different, the way that they experience depression or anxiety can hugely vary too.

Most people who experience mental health problems fully recover, or are able to live with and manage them – especially if they can get the right support at a young age. But different people with different mental health issues need different support – one size does not fit all when it comes to interventions. 70% of children and adolescents who experience mental health problems at an early age do not receive appropriate support and the generalised approach to diagnosis can affect patient trust in treatment.

In order to be able to support young people to prevent, effectively manage or overcome, their own mental health challenges, mental health practitioners need a better understanding of two things: the way that these disorders manifest through symptoms, and mechanisms of a mental health intervention that can help. Both of these areas have been explored through traditional research methods, but there is still a lot more that the mental health research community could learn about anxiety and depression.

One of the ways that research in these areas could be accelerated is through better use of data. The UK has a robust data ecosystem around mental health. There are opportunities to use existing mental health data to improve our understanding of anxiety and depression and to create new insights about the elements that make mental health interventions effective. Although there is a wealth of expertise across the UK mental health domain in terms of researchers and practitioners, support from people or organisations with a more comprehensive understanding of data may be required to unlock the most value from existing mental health datasets.

A Wellcome led Health Data Challenge Prize would support mental health experts to make better use of existing mental health data. The challenge would bring together a range of experts, from practitioners to researchers, and pair them with data scientists, to work together to improve the suitability and efficacy of mental health interventions for young people, aged 14 – 24.

The teams would have the option to focus on one of two challenge streams:

1. **Better understanding the role that individual symptoms** — the patterns or manifestations of anxiety and depression – play in the development of other symptoms of anxiety and depression.
2. **Better understanding the core components** – the active ingredients that make an intervention work, e.g. doing something on a daily basis – that make interventions for anxiety and depression work effectively.
Throughout the duration of the challenge, these multidisciplinary teams will partner with healthcare providers from the public, private and third sectors, as well as research organisations that collect data about mental health. With the provision of tailored training and support throughout the programme, this challenge also presents an opportunity to increase the long term data science capacity in the mental health space.

Consistent public engagement could lead to an increase in the trust that the public places in the use of health data to create positive outcomes, as well as healthcare and research organisations more broadly. The teams will work closely with young people to understand their needs and to build tools that help them to better manage their own mental health, or that help practitioners to make better decisions about the interventions that they prescribe and apply. Ultimately, the tools created from this challenge could help current and future generations of young people to better manage their own mental health, or to receive the right professional support when needed, leading to better mental health outcomes.

<table>
<thead>
<tr>
<th>Funding offered to participants*</th>
<th>Amount</th>
<th>Detail</th>
</tr>
</thead>
<tbody>
<tr>
<td>Seed funding for insights phase (6 months)</td>
<td>£500,000</td>
<td>This would amount to £25,000 for up to 10 groups per stream (core components and symptoms) to cover their time and expenses in taking part in the insights phase. This would be aimed at enabling researchers, schools and freelancers with relevant tech expertise to take part and contribute.</td>
</tr>
<tr>
<td>Seed funding for development phase (6 months)</td>
<td>£1,000,000</td>
<td>This would amount to £100,000 for up to 5 groups per stream (core components and symptoms). This would cover the development of the tools and the expansion of the teams to include requisite application development expertise.</td>
</tr>
<tr>
<td>Prize money</td>
<td>£1 million total</td>
<td>This would cover a prize of £200k for the winner and £100k for three runners up per stream (core components and symptoms). The money will need to be spent on rolling out and scaling the tool they have produced.</td>
</tr>
</tbody>
</table>

* Financial support will be appended with training and mentorship throughout.
** Complete budget can be found in Appendix 2.

Data science for reducing snakebite mortality: Summary

How can we use data science to reduce deaths from snakebite in India?

Farming rice in the South Indian state of Tamil Nadu poses many challenges. Poverty is already endemic, and the tropical region is on the forefront of the dangers of climate change. For all the threats to life and livelihood, there is perhaps one least understood by authorities but no less deadly to the rural populace. Anywhere between 3,000 and 10,000 people in Tamil Nadu will die from a snakebite this year, and deaths globally are on par with cervical and prostate cancer.
Working in fields recently converted from woodland to address rising food demand, without proper equipment, and often at night to escape the heat, farmers expose themselves to venomous snakes on a daily basis. Miles from the nearest medical clinic, and often one not stocked with the only real cure for snakebite, antivenom, victims often seek inadequate medical care, if at all.

Fortunately, the WHO recently classified snakebite as a ‘neglected tropical disease’, with the goal of 50% reduction in deaths and disabilities by 2030. However, data on the extent of the problem is lacking. Healthcare systems, strapped for funding, struggle to supply and allocate needed medicines and expertise in the right location and quantity. NGOs dedicated to prevention, education and healing, with equally limited resources seek better information on the shape of the problem in order to help tackle it.

A Health Data Challenge Prize run by Wellcome in Tamil Nadu would galvanise the research and data science community to create impactful solutions from existing but disparate data sources. The challenge would draw on the research insights from multiple disciplines, using data science to improve the epidemiology of snakebite. Over two years, global experts in the field of snakebite will work together with data scientists, ecologists, environmental scientists and others to build tools to combat the disease in Tamil Nadu. These challenge teams will access and combine datasets across healthcare, herpetology, climate, ecology and urbanisation. They will create and adapt predictive models to underpin tools that will help organisations working to effectively prevent and respond to snakebite.

These tools could help the healthcare system better plan a response according to the areas most at risk of snakebite, ensuring that the antivenom and medical expertise needed in rural areas are available. They could help NGOs plan their prevention and education better with this newly developed information. Importantly, the models created in Tamil Nadu could then be adapted to other regional contexts across southern Asia and Sub-Saharan Africa, meaning the Snakebite Health Data Challenge could provide a substantial benefit for communities at highest risk.

<table>
<thead>
<tr>
<th>Funding offered to participants*</th>
<th>Amount</th>
<th>Detail</th>
</tr>
</thead>
<tbody>
<tr>
<td>Seed funding for insights phase (6 months)</td>
<td>£250,000</td>
<td>This would amount to £25,000 for up to 10 groups to cover their time and expenses in taking part in the insights phase. This would be aimed at enabling snake catchers, healthcare professionals and researchers and tech experts to take part and contribute.</td>
</tr>
<tr>
<td>Seed funding for development phase (6 months)</td>
<td>£500,000</td>
<td>This would amount to £100,000 for up to 5 groups. This would cover the development of the tools and the expansion of the teams to include requisite application development expertise.</td>
</tr>
<tr>
<td>Prize money</td>
<td>£500,000</td>
<td>This would cover a prize of £200k for the winner and £100k for three runners up. The money will need to be spent on rolling out and scaling the tool they have produced.</td>
</tr>
</tbody>
</table>

* Financial support will be appended with training and mentorship throughout.
** Complete budget can be found in Appendix 2.
Data Science for improving antibiotic prescription and use: Summary

How can we use data science to optimise the use of antibiotics in order to minimise the local emergence and persistence of antimicrobial resistance in Uganda, Malawi and Kenya?

Antimicrobial resistance is one of the major public health concerns of this century. We all need to use antibiotics at some point in life, however antibiotic treatment is becoming less reliable as antimicrobial resistance (AMR) is evolving. Antibiotics have been a staple of modern medicine to fight illnesses, but antibiotic treatment has become less reliable as antimicrobial resistance (AMR) has evolved in many bacterial pathogens, causing the need for costly new treatments to be developed, while antimicrobial resistant organisms continue to take thousands of lives per year.\(^4\) If no action is taken the United Nations warn that drug-resistant diseases could cause 10 million deaths each year by 2050.

The World Health Organisation defines antimicrobial resistance as a phenomenon that happens when microorganisms (such as bacteria, viruses) change when they are exposed to antibiotics. As a result, the medicines become ineffective and infections persist in the body, increasing the risk of spread to others.

Using antibiotics that do not correspond to your condition – or a wrong combination of them, for the wrong period of time – builds resistance. Therefore, understanding how antibiotics are being prescribed and how patients are accessing them is key to understanding and stopping the spread of resistance.

In the UK, access to antibiotics is regulated, people need to visit a doctor and get a prescription before accessing antibiotics in a pharmacy. However, that is not the case everywhere. In some low and middle income countries, many antibiotics are available without prescription. They can be bought over the counter, in unlicensed drug stores and in open vans in markets, for example. People buy the antibiotics they can afford, take a lot of different ones at the same time, and not for the right amount of time. This builds a lot of resistance and is almost invisible as there is so little data on the prescription and consumption of antibiotics through these informal dispensaries. These issues have been observed in Uganda, Malawi and Kenya.

A Health Data Challenge Prize run by Wellcome could play a key role in motivating people with expertise in data science and public health to work together. Using behavioural and epidemiological data from a series of current and upcoming initiatives, the teams could map the flow of antibiotics through community pharmacies and into the population. These insights can then be used to increase understanding of the scale and impacts of resistance as much of the data about AMR comes from western sources.

Teams can then use these insights to create tools that support antibiotic use by giving the right information to decision makers for better interventions; to healthcare providers to encourage an appropriate use of antibiotics; and to NGOs to target campaigns to incentivise better

\(^4\) National Centre for Biotechnology Information, National Database of Antibiotic Resistant Organisms (NDARO)
prescription and use behaviours. If policymakers and healthcare providers can start working from a shared understanding of the scale and nature of antibiotic consumption, this has the potential to support real progress in the fight against antimicrobial resistance which is likely to become even more vital as COVID-19 continues to spread.

<table>
<thead>
<tr>
<th>Funding offered to participants*</th>
<th>Amount</th>
<th>Detail</th>
</tr>
</thead>
<tbody>
<tr>
<td>Seed funding for insights phase (6 months)</td>
<td>£250,000</td>
<td>This would amount to £25,000 for up to 10 groups too cover their time and expenses in taking part in the insights phase. This would be aimed at enabling snake catchers, healthcare professionals and researchers and tech experts to take part and contribute.</td>
</tr>
<tr>
<td>Seed funding for development phase (6 months)</td>
<td>£500,000</td>
<td>This would amount to £100,000 for up to 5 groups. This would cover the development of the tools and the expansion of the teams to include requisite application development expertise.</td>
</tr>
<tr>
<td>Prize money</td>
<td>£500,000</td>
<td>This would cover a prize of £200k for the winner and £100k for three runners up. The money will need to be spent on rolling out and scaling the tool they have produced.</td>
</tr>
</tbody>
</table>

* Financial support will be appended with training and mentorship throughout.
** Complete budget can be found in Appendix 2

Data Science for public health risks in urban environments:
Summary

How can we better understand and resolve public health risks in rapidly growing urban environments?

Increasing numbers of people living in towns and cities is one of the leading global trends of the 21st century that has a significant impact on health. Over 55% of the world’s population live in urban areas, a proportion that is expected to increase to 68% by 2050\(^6\). By 2030, projections indicate that two billion of the global urban population will live in slums, mostly in Africa and Asia \(^6\). However, the health of people who live in slums is a topic that has received little attention\(^7\).

A person’s risk of disease is affected by both personal factors, such as diet and their genetics as well as environmental factors such as sanitation and pollution, aka ‘neighbourhood effects’. Health authorities generally track information about individuals to inform decisions about public health and health services, but because of the high concentration of people in slum areas, environmental factors affect a far greater number of people at any one time. Information about

\(^6\) 68\% of the world population projected to live in urban areas by 2050, United Nations Department of Economic and Social Affairs, 2018
\(^7\) Slum health series, The Lancet, 2016
neighbourhood effects is therefore incredibly valuable for identifying highly impactful public interventions, but there is very little data about these communities.

For example, in many LMICs, open dumpsites are a common disposal method and in general, less than half of all solid waste in low-income countries is collected resulting in a significant amount of illegal and unsafe disposal\textsuperscript{8,9,10,11,12}. Poorly managed solid waste has health, environmental, and economic effects that multiply as waste accumulates. Uncollected solid waste increases exposure of all individuals in communities to vector-borne and zoonotic infectious diseases carried by birds, insects, and rodents. Over time, uncollected waste accumulates to block waterways, resulting in flooding, contaminated surface and groundwater, and emissions of greenhouse gases like methane. Abuja is one of the fastest growing cities in the sub-Saharan Africa but the city is said to lack the modern management techniques to meet the requirement of a rapidly expanding city\textsuperscript{13}, often resulting to major floods that lead to significant deaths due to a ‘lack of proper town planning, blocked waterways and poor drainage systems\textsuperscript{14}. Altogether, these neighborhood-level exposures lead to increased incidence of respiratory illness and diarrhea, and poorer mental health among individuals.

Municipal solid waste management is the largest budget item for city governments in most low-income and many middle-income countries and one of the largest employers\textsuperscript{15}. In LMICs, the amount of waste produced per person is expected to double in the next 20 years, and costs to manage solid waste will increase four to five fold\textsuperscript{16}. It is therefore in the interests of local government and policy makers in these regions to manage waste disposal better, but they have limited data about solid waste on which to base policies and allocate limited resources.

There is a thriving community of academics and data scientists using non-traditional data such as satellite imagery to develop tools and techniques to measure neighbourhood phenomena, for example the identification of slum area boundaries over time, or using machine learning to identify access ways and roads which can be used for specific public health risks such as solid waste management. However, this community rarely has health or medical backgrounds and therefore does not package or distribute data with health decision-makers in mind\textsuperscript{17}.

If Wellcome were to run a Data Science Challenge Prize which closely engaged with city authorities to understand their specific municipal challenges, and the communities themselves

\textsuperscript{8} Ziraba AK., Haregu, TN., & Mberu, B., A review and framework for understanding the potential impact of poor solid waste management on health in developing countries, Zirabaet al. Archives of Public Health (2016)
\textsuperscript{9} Adelowo OO, Akinlabi IA, Fagade OE. Environmental impact assessment of Attenda abattoir, Ogbomoso southwestern Nigeria on surface and groundwater quality using geo-electrical imaging and microbiological analysis. Environ Monit Assess. 2012;
\textsuperscript{10} Complex Urban Systems for Sustainability and Health research partnership
\textsuperscript{14} Nigeria floods kill more than 100, BBC, 2018, https://www.bbc.co.uk/news/world-africa-45546695
\textsuperscript{16} Ziraba AK., Haregu, TN., & Mberu, B., A review and framework for understanding the potential impact of poor solid waste management on health in developing countries, Ziraba et al. Archives of Public Health (2016)
to understand the effect of these challenges, it could motivate and support teams of data scientists working in the field of remote sensing to collaborate directly with public health experts. Through the combination of ‘remote sensing’ data such as satellite imagery, with administrative, municipal, health and community collected data, the challenge teams could develop models, tools and applications that support cities to make better decisions regarding public services that impact public health in urban areas. Due to the nature of ‘neighbourhood effects’, enabling cities to identify the specific interventions in slum areas that will have the biggest impact will support the health of a huge population, and help equip the authorities for dealing with their rapidly changing urban landscapes.

This research has identified that Nigeria and Kenya would be two suitable locations with a range of available data, networks of data scientists with sufficient capacity and initial interest from governmental authorities.

<table>
<thead>
<tr>
<th>Funding offered to participants*</th>
<th>Amount</th>
<th>Detail</th>
</tr>
</thead>
<tbody>
<tr>
<td>Seed funding for insights phase (6 months)</td>
<td>£250,000</td>
<td>This would amount to £25,000 for up to 10 groups too cover their time and expenses in taking part in the insights phase. This would be aimed at enabling snake catchers, healthcare professionals and researchers and tech experts to take part and contribute.</td>
</tr>
<tr>
<td>Seed funding for development phase (6 months)</td>
<td>£500,000</td>
<td>This would amount to £100,000 for up to 5 groups. This would cover the development of the tools and the expansion of the teams to include requisite application development expertise.</td>
</tr>
<tr>
<td>Prize money</td>
<td>£500,000</td>
<td>This would cover a prize of £200k for the winner and £100k for three runners up. The money will need to be spent on rolling out and scaling the tool they have produced.</td>
</tr>
</tbody>
</table>

* Financial support will be appended with training and mentorship throughout.
** Complete budget can be found in Appendix 2
Delivery Recommendations

Sequencing of the challenge prizes

Due to the impact of COVID-19, we recommend prioritising the mental health challenge to be delivered first. This is because it has significant relevance to the current situation in which lockdown, social distancing and economic impacts of the pandemic are likely to have a significant impact on the mental health of young people.

The snakebite challenge has been prioritised next as it is a very specific health challenge, focused on one location. It also is very closely aligned with the Snakebite Priority Area strategic objectives. While it will still require a setup phase in order to establish data access agreements and establish relationships with communities, stakeholders and delivery partners, there is a clear set of contacts to instigate this and a clear set of potential outcomes. Similarly, the DRI challenge aligns closely with the DRI priority area strategy, however it relies on a number of datasets which are planned to be collected over the next 18 months, which is why it has been sequenced later in the programme. The DRI programme could also be run in multiple locations due to the nature of the data, so there is some reflection to be done to identify the most suitable region to run this in. Finally, the urban health topic has many potential partners, has a strong connection to an existing data science community and aligns closely with the Our Planet Our Health priority area. However it is advised that this topic is reviewed in 2021 in order to ensure it aligns with the Environmental strategic pillar.

However, following the successful establishment of the mental health prize, we recommend re-evaluating the remaining three topics in light of the impact COVID-19 has on the relevant regions. It is likely the burden of disease may be much greater in countries within India and sub-saharan Africa so the topics may need to be rescoped or reprioritised (for example, there is an emerging link between COVID-19 and the use of antimicrobial drugs so it may be prudent to prioritise the DRI challenge before the snakebite challenge in order to respond to this issue). The strategic priorities within Wellcome may have also changed in that time, which may have an effect on some topics (for example, Wellcome is designing a pillar for Climate and Health, in addition to Our Planet Our Health, which may have an effect on the urban health topic).

Below is a summary of the potential sequencing of the challenges. This is under the assumption all four Challenge Prize topics would all be prioritised.

---

Also available in Appendix 10
### Challenge Prize Programme Timeline

<table>
<thead>
<tr>
<th>Year 1 (2020-2021)</th>
<th>Year 2 (2021-2022)</th>
<th>Year 3 (2022-2023)</th>
<th>Year 4 (2023-2024)</th>
<th>Year 5 (2024-2025)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Suggested Challenge Topic</strong></td>
<td><strong>Programme Management &amp; Evaluation</strong></td>
<td><strong>Programme Management &amp; Evaluation</strong></td>
<td><strong>Programme Management &amp; Evaluation</strong></td>
<td><strong>Programme Management &amp; Evaluation</strong></td>
</tr>
<tr>
<td>Q3</td>
<td>Q4</td>
<td>Q1</td>
<td>Q2</td>
<td>Q3</td>
</tr>
<tr>
<td><strong>Mental health for young people (UK)</strong></td>
<td>Programme Set Up</td>
<td>Model development phase + demo day</td>
<td>Tool development phase and prize</td>
<td>Prize Comms, Evaluation &amp; Sustainability</td>
</tr>
<tr>
<td><strong>Scalable (India (Tamil Nadu))</strong></td>
<td>Set up and recruitment</td>
<td>Model development phase + demo day</td>
<td>Tool development phase and prize</td>
<td>Prize Comms, Evaluation &amp; Sustainability</td>
</tr>
<tr>
<td><strong>Drug Resistant Infections (Kenya, Malawi or Uganda)</strong></td>
<td>Set up and recruitment</td>
<td>Model development phase + demo day</td>
<td>Tool development phase and prize</td>
<td>Prize Comms, Evaluation &amp; Sustainability</td>
</tr>
<tr>
<td><strong>Remote Sensing for Urban Health (Kenya or Nigeria)</strong></td>
<td>Set up and recruitment</td>
<td>Model development phase + demo day</td>
<td>Tool development phase and prize</td>
<td>Prize Comms, Evaluation &amp; Sustainability</td>
</tr>
</tbody>
</table>

*Reflection (Insights applied to subsequent challenge prizes)*

*Evaluation (Evaluation of success of prizes)
Summary of proposed phases

The following descriptions are generalised and are subject to slight alterations depending on the requirements of the specific challenge prize and topic.

<table>
<thead>
<tr>
<th>Phase</th>
<th>October</th>
<th>November</th>
<th>December</th>
<th>January</th>
<th>February</th>
<th>March</th>
<th>April</th>
<th>May</th>
<th>June</th>
<th>July</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reflection &amp; Evaluation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Set up and recruitment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Launch &amp; Applications</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Insights phase</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Development phase</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Judging Period</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sustainability Period</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1 - Set up + recruitment phase
Up to 9 months for building relationships with stakeholders, delivery partners and data partners. This phase will require the topic to be verified and refined with these stakeholders and delivery plan to be finalised, followed by recruitment for expressions of interest.

2 - Launch Lab + Application phases
Multi day launch event to introduce the prize and support people to form teams. Team applications within 4 weeks.

3 - Model development phase + demo day
This phase will be between 6-9 months and focus on training and the development of data insights. The teams will be encouraged to focus on engaging with the available data and modeling it to develop insights. During this time they will receive some seed funding (up to £25k for 10 groups), data access support, training and mentorship. At the end of the phase teams demo their work and the top teams are selected for the next phase.

4 - Tool development phase
This phase will run for between 6-9 months. These teams will receive further seed funding (£100,000 for up to 5 groups) to enable them to develop the tool or application outlined in the insights phase. Further training and mentorship will be available including business strategy support and implementation planning.

5 - Judging
The development phase will culminate in a public demo day attended by as many media representatives and senior stakeholders attending as possible. Each of the teams present their work and a panel of leading representatives in the field will select a winning team who will receive £200k and three runners up who will receive £100k each. This money will be explicitly for the implementation and scaling of the tool they have developed.

6 - Sustainability Phase
This phase of up to 9 months is aimed at enabling the winning teams to implement their innovations with beneficiaries using the funds from the prizes. They shall also receive further support from the delivery teams to make introductions to would–be users, investors,
commissioners and others, checking in on progress and providing ad hoc advice from mentors, feedback from competitors, judges and partners

7 - Comms & Evaluation
At the end of the judging and sustainability phase there shall be independent evaluations made of the prize as a whole drawing on reporting material and interviews with participants and stakeholders. The interviews will also be used to develop case studies of each output and prize in order to share learnings and successes for others interested in the field of health data innovation. More detail on the evaluation structures can be found in the following Communications Strategy\(^{19}\) and the Governance & Evaluation sections.

Governance, Reporting & Evaluation

Governance
For the challenge programme to be successful, a management structure will need to be put in place in order to oversee the programme as a whole and the individual challenge prizes. Wellcome’s Data for Health and Science team are the priority stakeholders, in partnership with the relevant subject specific priority area (Mental Health, Snakebite, Drug Resistant Infections and Our Planet Our Health). Governance structures should be built around their needs and processes but as an overall structure we recommend having a tiered approach with governance and advisory boards at both the programme and prize levels.

The prize governance boards will be the bodies that the prize delivery teams report to within Wellcome. They will meet quarterly and be responsible for the delivery of that specific prize. The members of the prize governance board will include Wellcome stakeholders from the Data for Science and Health Team along with stakeholders from the relevant topic area and key stakeholders from the Challenge itself such as representatives from the Ministry of Health.

The programme governance board will be the body that the programme delivery team will report into. They will be responsible for the operational delivery of the programme. They will meet every six months and consist of key stakeholders within Wellcome across the Data for Science and Health team and a subject specific representatives from each of the prize governance boards.

The prize advisory boards will be formed of leading experts in the field and will inform the key decisions of the prize. Judges for the prizes may come from the advisory board and they will provide invaluable communications and dissemination support. They will meet at the start of the prize in order to inform the set up phase, mid way through the prize, and at the end (roughly every 6 months).

The programme advisory board will be formed of experts across the health and data innovation field in order to inform key programmatic decisions. They will meet annually.

The programme delivery team shall be responsible for:

\(^{19}\) Full strategy can be found in Appendix 4.
● Programme project management: Creating and running delivery and governance frameworks, managing the set up, administration, reporting and delivery of the programme as a whole across the five year period. Managing procurement of all subcontract support.
● Financial management: Budget management and accounting, reporting regularly on drawdown and expenditure.
● Partner relationship management: Owning and maintaining relationships and stakeholders across the challenges, including prize specific collaborators and delivery partners
● Coordinating programme wide legal and comms support: Legal support will be needed to set up the legal framework for the challenges in terms of template data sharing agreements, participation contracts and IP advice. Comms will be needed to create the comms assets across the full prize programme, including brand, digital assets such as website and social media channels which can be used and adapted across each prize, and press and marketing in order to promote the impact of all the prizes to the global health sector. The full strategy is detailed below.
● Running reviews and evaluation: Monitoring the progress of the challenges and making changes to the delivery model as necessary to ensure it has the widest impact.
● Revisiting and reviewing the feasibility of later challenge prize topic: Due to COVID-19 and evolving Wellcome priorities in the environmental health field, it is important to revisit the challenge topics scheduled for later in the programme in order to re-verify their feasibility.

The prize delivery teams will work collaboratively with the programme delivery team and governance and advisory boards to develop and deliver detailed delivery plans for individual Challenges which will then be approved by Wellcome and reported on. These delivery plans will include deliverables including (but not limited to):
● Eligibility criteria for participants
● Defining and delivering the approach to the challenge set up phase
● Individual budgets and delivery plans
● Sustainability scoping and planning
● Legal oversight
● Stakeholder engagement
● Prize specific communications
● Recruitment
● Data access establishment
● Event planning and management
● Training and mentorship coordination
● Judging and prize giving

Reporting
The prize delivery teams shall submit monthly written reports on progress, financial status and risk mitigation to the governance board which shall meet every quarter to review progress. These monthly reports will be supported by the tracking of each of the participating teams in accordance with the KPIs set out in the theory of change and delivery plans. The advisory boards shall also receive these reports.
The programme delivery team shall similarly prepare monthly reports which are submitted to the programme governance board which compiles the key points across the various prizes to give the programme governance board a comprehensive overview of all activity.

In parallel there is a quarterly meeting with representatives from each of the delivery teams across the prizes to share progress, updates, learnings from the reflection points, blockers and risks.

The diagram below shows the relationship between the various delivery teams, governance and advisory boards, reports and meetings.

Annual events

Annual events which will start midway through the first challenge prize, and then occur annually after that until the end of the programme. These events will publicly showcase the activities of all the prizes together. They will be attended by the delivery teams, governance boards, advisory boards, and key stakeholders. Participants are invited to showcase their work. This builds networks across the prizes, and creates international media opportunities. Their timing can be seen on the programme sequencing document.

*NB: Scheduling the formal meetings for the Programme Advisory Board six months after the annual events will result in their engagement twice a year.*
Evaluation

Evaluation will happen at both the programme and prize level and will occur on an ongoing basis in order to apply the insights to subsequent prizes. In order to ensure rigor and impartiality, the evaluation framework will be developed in collaboration with an independent third party who will subsequently complete the final evaluation of the prizes and their outputs once they have created their tools. This is why there is £500,000 in the budget for evaluation.

KPIs will be measured using qualitative and quantitative metrics. Specific KPIs for both programme and prize level will be decided on as part of the first governance board meeting in alignment with the programme and prize theories of change.

The delivery teams themselves will monitor the actual KPIs through monthly reporting and enable iteration across the prizes through the reflection points.

In order to ensure that learnings are transferred from one prize to another, reflection points have been built in across the delivery of the prizes. Learnings and insights from these can be shared by the prize delivery teams in the quarterly meeting across all prize teams. The diagram below shows the relationship between the programme and prize level evaluations along with the iteration across prizes.
• **Programme Theory of Change:** This articulates the overall context for the challenge prizes and links the consistent activities undertaken with the impacts the prize programme is looking to achieve. This is used as a framework to create the prize specific theories of change, delivery plans and evaluation frameworks.

• **Reflection points:** Reflection points will occur at the end of every key milestone (set up and recruitment, model development phase and demo day, tool development phase and prize). These provide an opportunity to iterate based on learnings, pivot the direction of the programme and re-evaluate the sequencing and focus areas of the Challenge Prizes.

• **Output reviews:** following the delivery of the outputs, they will have case studies written about them, which will then be revisited 6 months to one year after the end of the prize with a view to assessing the longer term uptake/embedding of those outputs.

• **Challenge prize evaluation** At the start and end of each prize, the evaluation framework will be defined in alignment with the theory of change and delivery plan in collaboration with an independent third party. At both the point when the prizes are awarded, and following the sustainability phase the third party evaluator will complete an initial and final evaluation of the prize drawing on the monthly reports, documentation from the reflection points and interviews with the participants and beneficiaries. These interviews will form case studies.

**Programme risk register**

A risk register is a key tool for documenting risks, and mitigation processes to manage each risk. It is essential to the successful management of a challenge programme, as risks might arise at different points and for various reasons.

Some risks have already been identified in the Wellcome scoping risk register\(^{20}\) across the programme as a whole and the specific topics. This document is to be considered as a live document, it would be updated throughout running the challenge programme. Risks would be progressively logged on the register and actions would be taken to respond to each risk.

Key risks across the programme include the lack of existing data infrastructure in many sub-Saharan African countries, the lack of interoperability for existing data sets which will need significant cleaning, and the need to ensure that the beneficiaries of the challenges are effectively engaged all the way through the programme to ensure that the innovations created through the challenges are actually implemented and used. Without this last factor being taken into consideration, the programme risks creating innovations which work as demonstrators but have limited actual impact.

\(^{20}\) Appendix 3
Programme Budget

A full budget estimate can be found below and separately in Appendix 2 detailing estimated cost across the individual challenge prizes and programme coordination. This budget includes detailed annotations as to the rationale for different forecasted costs. They take into consideration the individual needs of the different challenges and can be used to inform requirements listed on an ITT.

The costs have been benchmarked against other similar programmes including Data Pitch and the Open Data Challenge Series.\(^2\)

The costs for the Mental Health challenge are higher as the scoping has identified two distinct sub-topics which participants can choose to engage with. It is therefore twice the size of the other challenge prizes. Costs will of course vary based on different currencies, and the scale of the prize undertaken. For example, the delivery team could choose to increase or decrease the amount of participants and associated seed funding.

<table>
<thead>
<tr>
<th></th>
<th>Mental Health Challenge Prize</th>
<th>Snakebite Challenge Prize</th>
<th>DRI Challenge Prize</th>
<th>Urban Health Challenge Prize</th>
<th>Programme costs</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Time</strong></td>
<td>£580,400</td>
<td>£755,400</td>
<td>£605,400</td>
<td>£705,400</td>
<td>£1,542,800</td>
</tr>
<tr>
<td><strong>Expenses</strong></td>
<td>£46,200</td>
<td>£77,100</td>
<td>£72,100</td>
<td>£87,100</td>
<td>£300,000</td>
</tr>
<tr>
<td><strong>Participant Funding</strong></td>
<td>£2,500,000</td>
<td>£1,250,000</td>
<td>£1,250,000</td>
<td>£1,250,000</td>
<td>£0</td>
</tr>
<tr>
<td><strong>Challenge Total</strong></td>
<td>£3,126,600</td>
<td>£2,082,500</td>
<td>£1,927,500</td>
<td>£2,042,500</td>
<td>£1,842,800</td>
</tr>
</tbody>
</table>

Programme total  £10,969,900  
Total Time  £3,609,000  
Total Expenses  £536,300  
Total Funding  £3,750,000  

\(^2\) Details of which can be found in Appendix 8
<table>
<thead>
<tr>
<th>Item</th>
<th>Description</th>
<th>Cost Type</th>
<th>Cost Notes</th>
<th>Mental Health</th>
<th>Cost Notes</th>
<th>Scientific</th>
<th>Cost Notes</th>
<th>ESR</th>
<th>Cost Notes</th>
<th>Urban Health</th>
<th>Cost Notes</th>
<th>Programme wide</th>
<th>Cost Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Total Funding**

- **Grand total**: £1,250,000
- **Prize money**: £200,000
- **Judging event expenses**: £25,000
- **Stakeholder expenses**: £40,000
- **Data QA access control**: £70,000
- **Prize specific delivery support**: £50,000
- **Branding support management**: £60,000
- **Financial management**: £50,000
- **Estimated Programme Budget**: £10,969,900

**Funding Challenges**

- **£1,250,000**: 6 days a month for £500 per day
- **£200,000**: 4 days a month across 6 months at an average cost of £800 per day
- **£25,000**: 8 days for one month at an average cost of £800 per day
- **£40,000**: 12 days a month, for 60 months across 5 years at an average cost of £800 per day
- **£70,000**: 5 days a month across 9 months at an average cost of £800 per day
- **£50,000**: 5 days a month across 6 months at an average cost of £800 per day
- **£50,000**: 3 days a month across 60 months across 5 years at an average cost of £800 per day
- **£800,000**: 8 days for one month at an average cost of £1000 per day
- **£250,000**: 5 days a month for £500 per day

**Time**

- **£200,000**: 5 days a month across 9 months at an average cost of £800 per day
- **£25,000**: 12 days a month for 60 months across 5 years at an average cost of £800 per day
- **£40,000**: 12 days a month, for 60 months across 5 years at an average cost of £800 per day
- **£70,000**: 5 days a month across 9 months at an average cost of £800 per day
- **£50,000**: 5 days a month across 6 months at an average cost of £800 per day
- **£50,000**: 3 days a month across 60 months across 5 years at an average cost of £800 per day
- **£800,000**: 8 days for one month at an average cost of £1000 per day
- **£250,000**: 5 days a month for £500 per day

**Cost**

- **£19,200**: 4 days a month across 6 months at an average cost of £800 per day
- **£6,400**: 8 days for one month at an average cost of £800 per day
- **£2,082,500**: 8 days for one month at an average cost of £800 per day
- **£1,250,000**: 6 days a month for £500 per day
- **£200,000**: 4 days a month across 6 months at an average cost of £800 per day
- **£25,000**: 8 days for one month at an average cost of £800 per day
- **£40,000**: 12 days a month, for 60 months across 5 years at an average cost of £800 per day
- **£70,000**: 5 days a month across 9 months at an average cost of £800 per day
- **£50,000**: 5 days a month across 6 months at an average cost of £800 per day
- **£50,000**: 3 days a month across 60 months across 5 years at an average cost of £800 per day
- **£800,000**: 8 days for one month at an average cost of £1000 per day
- **£250,000**: 5 days a month for £500 per day

**Notes**

- **£19,200**: 4 days a month across 6 months at an average cost of £800 per day
- **£6,400**: 8 days for one month at an average cost of £800 per day
- **£2,082,500**: 8 days for one month at an average cost of £800 per day
- **£1,250,000**: 6 days a month for £500 per day
- **£200,000**: 4 days a month across 6 months at an average cost of £800 per day
- **£25,000**: 8 days for one month at an average cost of £800 per day
- **£40,000**: 12 days a month, for 60 months across 5 years at an average cost of £800 per day
- **£70,000**: 5 days a month across 9 months at an average cost of £800 per day
- **£50,000**: 5 days a month across 6 months at an average cost of £800 per day
- **£50,000**: 3 days a month across 60 months across 5 years at an average cost of £800 per day
- **£800,000**: 8 days for one month at an average cost of £1000 per day
- **£250,000**: 5 days a month for £500 per day
Communications Strategy Summary

The full communications strategy can be found in the Appendix 4.

The Wellcome Health Data Challenge programme is being developed in the context of Covid-19, where public health, healthcare and health data are under scrutiny. It will be vital for the programme to have a strong, relevant, public voice, and maximise its unique value, in order to attract the participation and attention it needs to be successful.

In order to achieve as much interest and participation as possible, communications messaging needs to be strong, clear and to the point, appealing to the audiences and stakeholders we are aiming to engage and motivate them to get involved.

This document is our Communications strategy, researched and developed to attract the maximum number of high quality applicants to each of the Challenges. The building blocks of our strategy are focused on a strong visual brand and an attractive, interactive website that engages the user.

As part of this work, we have: reviewed other health challenges across the world, held a number of 1-2-1 fact-finding interviews with international experts and conducted SWOT and PESTLE analyses. The work has informed this paper at every level. In it, we present our findings and recommendations for a communications approach to the Wellcome Health Data Challenge programme.

In developing this strategy, we aimed to identify the specific communication needs of the Health Data Challenge programme and the participating communities. We asked, “what do we need to help this initiative to succeed?”

The approach set out in the strategy document (Appendix 4) will lead to communication activities that:

- Broaden awareness and engagement amongst a range of stakeholders and influencers
- Drive high-calibre, high-volume entries
- Ensure the programme is unique and fosters ethical data innovation to solve global health problems

The aim is for the Challenge programme to be promoted within the relevant countries so that quality entries can be created and submitted, ready for judging and onboarding.

The communications and marketing work set out in this plan will support that aim by communicating the important role the Challenges will play in improving healthcare across the world.

- **Develop programme branding** - A set of branding should be developed and used in all external and internal communications. Our research into the most successful, similar programmes revealed that they all had strong visual identities that help the audience understand the project as a whole, capturing and holding the audience’s attention. It is therefore recommended that written content is kept to a minimum with the main focus being on auditory and visual information (e.g. videos).
• **Launch the Challenge programme website**- The introduction of a dedicated website is intrinsic to the success of this programme. It is the primary channel that will embed the branding and in turn, successfully promote the programme’s aims and objectives resulting in high-quality applications that fit the brief. A Challenge programme’s website should be developed to include: a visual timeline of the Challenge programme; written and video explainers of the programme’s vision, aims and objectives; case studies from the affected communities; a resources page featuring toolkits and downloadable material

• **Develop a Toolkit**- This should be a one-stop-shop containing information and content for all audiences

• **Using judges’ network**- To help communicate the programme more widely, it is recommended that the expert judges’ networks are leveraged

• **Host online webinars**- these online information clinics should be designed to provide potential applicants with the chance to ask questions about the application process and raise any issues they would like to discuss

• **Develop a mobile application**- This should allow applicants to track their application and get updates on the go

• **A virtual, global launch event that ‘Chases the Sun’**- Launching the programme in key geographies, around the world in multiple time zones, to raise interest and encourage participation.

• **Annual events to showcase the work happening across all the prizes**- These public events will start mid-way through the first challenge prize and then take annually thereafter. They will bring together the delivery teams, governance boards, advisory boards, participants, press and key stakeholders to showcase the activities of all the prizes collectively. These events will build networks across the prizes, and create international media opportunities to promote the work.
Mental health data challenge: Detailed Scoping

Summary Recommendation

Key challenge question
How can we support the development of more effective interventions for addressing anxiety and depression in young people and adolescents (ages 14-24)?

Why?

75% of mental health problems start before the age of 18, with consequences that are often lifelong.22 Around 14% of adolescents between the ages of 13-17 experience a depressive episode and nearly one in three adolescents meet the criteria for an anxiety disorder before the age of 18.23 Supporting young people to take care of their mental health and wellbeing could have a largely positive impact on the economy and society, as research suggests that childhood mental health problems can lead to substantial additional costs in adulthood, in areas such as criminal justice, education and health.24 However, up to 70% of children and young people that display early symptoms of anxiety and depression, or have been diagnosed with an anxiety or depressive disorder, do not receive appropriate support.25

While many anxiety and depression treatment approaches have proven to be effective for some people, these interventions will not always work for everyone. There could be many reasons as to why individuals respond positively to some interventions, and less so to others, but there is a real lack of compelling evidence to help mental health practitioners understand what these reasons are.26 In order to be able to make better decisions about which interventions young people with an anxiety or depressive disorder need, mental health practitioners first need to better understand the conditions themselves.

Existing diagnostic guides such as the Diagnostic and Statistical Manual of Mental Disorders (DSM–5)27 can be unreliable, as many of the symptoms that form the criteria for one condition also overlap with several other conditions as well. It is estimated that 45% of people with a particular mental health condition also satisfy the criteria for two or more other conditions.28 We also know very little about the underlying mechanisms of existing mental health and wellbeing interventions, which makes it hard to determine which ‘core components’ - the active

---

ingredients that make an intervention effective, such as the regularity of the intervention - would be useful across different types of mental health interventions. The lack of understanding about core components of interventions can make it difficult to improve existing interventions, or to create new ones.

What?
The UK has a robust data ecosystem around mental health presents opportunities to use existing mental health data to improve our understanding of anxiety and depression and to create new insights about the core components of mental health interventions. Although there is a wealth of expertise across the UK mental health domain in terms of researchers and practitioners, support from people or organisations with a more comprehensive understanding of data may be required to unlock the most value from existing mental health datasets.

We propose a data science challenge prize that supports multidisciplinary teams of data scientists, as well as mental health researchers and practitioners, to examine how to make mental health interventions more effective. The challenge would focus on creating insights and tools to help mental health practitioners, such as clinicians, to make better decisions about interventions for young people (ages 14-24) who are at risk of developing, or have already been diagnosed with, an anxiety or depressive disorder. It would include a package of seed funding, training for all participants and sustainability support for the best three solutions worth £2.5million.

The challenge would focus on two separate streams:

1. **Better understanding the role that individual symptoms play in the development of other symptoms of anxiety and depression**, which could help clinicians, researchers and individuals to improve existing, or develop new, mental health interventions which address those symptoms. By ‘symptoms’ we mean the physical, cognitive, emotional and behavioural patterns or manifestations of anxiety and depression.

   There are many different types of data that would be useful in better understanding the symptoms of anxiety and depression. Some examples of data sources for key types of symptoms include:
   a. Physical: e.g. trouble sleeping
   b. Behavioural: e.g. lack of appetite
   c. Emotional: e.g. persistent low mood
   d. Cognitive: e.g. persistent negative thoughts
   e. Social: e.g. avoiding social activities

2. **Better understanding the core components that make interventions for anxiety and depression work effectively**, which could help clinicians, researchers and individuals create better interventions for young people who are at risk of developing, or have been diagnosed with, an anxiety or depressive disorder. By ‘core components’, we mean the active ingredients that make an intervention work, e.g. doing something on a daily basis.
An interventions focused challenge would seek to use data about traditional clinical interventions, digital interventions and mental health promotion and wellbeing interventions, from multiple sources.

Where?
This document includes information on the UK, but mostly focuses on England, Wales and Scotland. The UK has been prioritised as the mental health data ecosystem is relatively strong and there is a lot of potential to create positive impact through a data science challenge.

There may be opportunities to scale the activities from one country to another, adjusting for specific needs and stakeholders, but this has not been scoped as part of this work.

Who?

| Creators (Participants) | • Multidisciplinary teams with expertise in data science, mental health research and practice (data scientists, clinicians, researchers) could identify new insights and develop new solutions (analysis, products, tools) based on data combined for the first time.  
• Young people (ages 14-24) with lived experience of anxiety and depression will play a core role in shaping the development of tools and products.  
• This expertise will be supported by people with skills in developing tools and services. |
| --- | --- |
| Decision makers and potential tools created to aid them | • Tools and products established that enable clinicians, other mental health care providers and young people themselves to make better decisions about the right treatment approaches for anxiety and depression in young people and adolescents.  
• This can include tools that provide recommendations for interventions, approaches and techniques for young people and adolescents in general as well as tools to support decision-making for individual patients.  
• Primary decisions makers are likely to be mental health practitioners, such as clinicians, or young people with lived experience of anxiety and depression.  
• Tools and products are likely to be beneficial to researchers.  
• Secondary decision makers might include funders of mental health research and practice, policy makers at a national and local level as well as regulatory bodies. |
| Beneficiaries | • Mental health practitioners who are responsible for supporting young people with mental health problems  
• Young people (ages 14-24) with lived experience of anxiety and / or depression, who need support to |
better manage their symptoms

Background Information

According to the 2013 Chief Medical Officer’s report, mental health problems cost the UK between £70-100 billion per year. Mental health research in the UK receives a comparatively small amount of funding (£125 million a year) compared to other major health research areas, such as cancer research (£404 million a year). This may be, in part, due to how reliant the mental health space is on government grant funding, as 67% of mental health research is funded by the UK Government, with around 30% being provided by other funders, such as philanthropic organisations, and just under 3% coming from public donations collected by charitable organisations. 

Evidence from as early as 1998 shows that poor mental health at a young age can have many negative long term consequences, such as social exclusion, an inability to maintain interpersonal relationships, anti-social or criminal behaviour and issues with long term employment. In the short term, understanding and addressing anxiety and depression at the earliest possible stage could mean that children and young people are protected from these longer term consequences. Despite these findings, research on children and young people’s mental health remains relatively underfunded, with only 30% of UK mental health funding invested in this area. In order to understand how to best support young people if they are developing symptoms of anxiety and depression, more effort should be invested in understanding these symptoms, and also understanding what makes a mental health intervention effective at reducing the effects of these symptoms, or potentially eliminating them completely.

Almost half of UK mental health funding is spent on underpinning research, which looks at the structure and function of the brain and how it works, or aetiology, which tries to understand the factors that lead to the development of mental health conditions. These research areas provide an important foundation for other research into treatments, interventions and the management of mental health. There is a plethora of existing information available on what the symptoms of anxiety and depression are, through sources such as the NHS, Mind, Rethink and more, as well as plenty of research papers and studies that try to examine particular symptoms. Much of this research assumes that depression and anxiety symptoms are interchangeable factors that predict a specific underlying mental health condition, but this type of assumption makes it difficult to explain why there are so many variations of anxiety and depressive disorders, and why some people who are diagnosed with a particular condition are more resistant to a specific

---

type of intervention than someone else with the same diagnosis. Recent research\(^{35}\) suggest that symptoms of anxiety and depression could be thought of as being part of a network of symptoms, which could explain why there are variables between individual cases of anxiety and depression. Understanding the role that each individual symptom plays in the wider network of symptoms of a person’s mental health condition could help practitioners to create more targeted interventions that address the most problematic symptoms.

There are also a number of tools and tests that mental health practitioners can use as an aid for understanding how symptoms manifest as an anxiety or depressive disorder. For example, the DSM–5\(^{36}\) has historically been used by psychiatrists as a definitive list of all recognised mental health conditions, and their associated symptoms. However, the DSM-5 is also considered to be fairly flawed, in that the pharmaceutical industry has a strong influence on revisions that are made and there has been a gradual shift towards considering some fairly normal patterns of behaviour or mood as abnormal.\(^{37}\) Without reliable tools to inform how they make decisions, practitioners are likely to make mistakes in diagnosing someone with an anxiety and depressive disorder,\(^{38}\) which can lead to further destabilisation of the individual’s condition, and prescription of inadequate or inappropriate treatment that will not help with the symptoms that the person is experiencing.\(^{39}\)

Evaluating specific interventions is a common research topic in the mental health space. Although there are many studies that report mental health interventions to be successful with a particular participant pool, it is widely accepted that the same mental health interventions are not effective for every single person with a particular condition. Research acknowledges that some people respond more positively to interventions than others, but results are often inconclusive as to why.\(^ {40}\) There is not much existing research in this area, but Wellcome’s mental health team has recently published an opportunity for researchers to examine potential core components, by reviewing existing interventions and providing an insight analysis for one core component that they think is the most promising. Once a stronger evidence base for the core components has been established, it is important that these insights are used to create tools and applications that help mental health practitioners to implement the findings into the interventions that they are prescribing or administering to young people.

The UK has an enormous data landscape around mental health. There are opportunities to use existing mental health data to improve our understanding of anxiety and depression and to create new insights about the core components of mental health interventions. Although there is a wealth of expertise across the UK mental health domain in terms of researchers and practitioners, support from people or organisations with a more comprehensive understanding of data may be required to unlock the most value from existing mental health datasets.

---


Most of the stakeholders that we engaged with are in favour of the use of data science in mental health research. The data providers that we engaged with who are also healthcare providers, such as Place2Be, are more open to working with data scientists as they feel that they lack the skills to get the best value from the data they have.

There are some existing networks of mental health data scientists, such as the Mental Health Data Science Group which is convened by MQ Mental Health. However, there is not an abundance of people with data science skills in the mental health space. Many of the researchers that were engaged during the scoping process were either working with data scientists from other academic disciplines, such as mathematicians, or did not have access to a colleague with a data science background.

There is interest from the data science community - both in academic and commercial circles - to become more involved in the mental health space, because it is a relatively unexplored area, with lots of interesting challenges to be solved. A challenge structure would be a great opportunity to enable people with data science expertise to connect with people who have mental health expertise and organisations that steward mental health data.

**Theory of change**

A full theory of change is available in Appendix 1. The underlying hypotheses for the two streams of the challenge are:

**Stream 1:** Better understanding the core components that make interventions for anxiety and depression work effectively

- **If we:** run a data science challenge which brings together teams of data scientists and mental health researchers to use data about the experiences of anxiety and depression in young people (ages 14-24) from multiple sources
- **Then:** this will enable those teams to use insights about the core components of effective interventions for anxiety and depression to create tools and products
- **Which:** can be used by mental health practitioners and young people to inform the creation of more effective intervention and prevention approaches
- **Ultimately leading to:** better and more tailored provision of care or self-care activities for young people with anxiety and depression

**Stream 2:** Better understanding the role that individual symptoms play in the development of other symptoms of anxiety and depression

- **If we:** run a data science challenge which brings together teams of data scientists and mental health researchers to use data about the experiences of anxiety and depression in young people (ages 14-24) from multiple sources
- **Then:** this will enable those teams to use insights about the role that individual symptoms play in the development of other symptoms of anxiety and depression to create tools and products
- **Which:** can be used by mental health practitioners and young people to inform the creation of more suitable intervention and prevention approaches
- **Ultimately leading to:** better and more tailored provision of care or self-care activities for young people with anxiety and depression
Likely types of benefit

The challenge prize aims at developing new insights about interventions, approaches and techniques for addressing anxiety and depression in young people.

The challenge proposes two primary avenues for doing this:

a) **Better understanding of anxiety and depression** in young people and adolescents: Exploring the role individual symptoms play in the development of other symptoms of depression and anxiety, and whether this changes over time, and creating insights about the symptoms that should be prioritised in interventions. The tools and products created from this challenge will help clinicians, other mental health care providers and individuals adopt more suitable intervention models.

b) **Providing better care** to young people and adolescents: Improving our understanding of the core components that make interventions for anxiety and depression work effectively and that should be integrated consistently in various types of interventions. The tools and products created from this challenge will help clinicians, other mental health care providers and individuals create better interventions, including the actions taken by young people themselves to manage their condition.

The solutions to the challenge prize might include for instance preliminary lists of core components for key types of interventions addressing anxiety and depression (e.g. traditional, digital, mental health promotion and well-being interventions) and network structures depicting the role of individual symptoms in the development and persistence of other symptoms of anxiety and depression and the strength of the relationships between symptoms. Other outputs that facilitate the application of the insights from the challenge into clinical practice will be encouraged.

Potential time to impact

Provided access to sufficient data, it may be possible to achieve impact with solutions within a medium to long term (two to five years) following the end of the prize. Efforts to build solid long-term relationships between data scientists, clinicians, mental health researchers and young people and adolescents that are based on trust will be essential for the success of the challenge initiative.
Success indicators

**Impact indicator:** Improved mental health outcomes among young people and adolescents in the UK due to more efficient and suitable treatments for anxiety and depression.

- This impact would need to be based on studies comparing the efficacy of any new treatments with treatment as usual, as it would be difficult to see and measure a whole population change in the space of five years.

**Mid-term outcome indicator:** Use of products and tools created during the challenge to inform decision-making in clinical practice and in digital intervention solutions offered to young people and adolescents in the UK.

Additionally to these success indicators, this challenge prize will meet Wellcome’s intended aims for the challenge prizes in the following ways:

- **Increasing public trust:** young people will be encouraged to take part of the co-design process of the challenge, ensuring that their needs are taken into account alongside the needs of mental health practitioners. Representatives will also be part of the judging panels.

- **Increasing capacity:** Research suggests\(^{41}\) there is considerable value in bringing together mental health researchers and data science experts. Training and mentorship will happen throughout the data challenge programme, in order to help build long term data science capacity in the mental health space.

Strategic alignment with Wellcome

Wellcome’s mental health priority area has an [ongoing commitment](#) to supporting research into mental health and has recently committed £200million in funding towards improving interventions for anxiety and depression in young people aged 14-24. The Wellcome team has very recently published a commission for organisations to explore core components, which is an area that this challenge would aim to explore.

This proposal aligns very closely with Wellcome’s [mental health programme strategy](#). As a result of this alignment, the Wellcome mental health team have expressed an interest in co-funding this challenge prize.

---

\(^{41}\) MQ Mental Health. 7 ways data science is transforming mental health research. [https://www.mqmentalhealth.org/news-blog/post/7-ways-data-science-is-transforming-mental-health-research](https://www.mqmentalhealth.org/news-blog/post/7-ways-data-science-is-transforming-mental-health-research)
Data landscape summary

Data stewards in the mental health space
In the UK, there is a wealth of information being generated about young people’s mental health by a variety of data stewards. The majority of this data is being collected by healthcare providers, academic researchers and commercial organisations.

The primary provider for mental health services in the UK is the NHS, which invests around £677 million a year into mental health services for children and young people.\(^\text{42}\) Primary care services for mental health are usually provided by general practitioners, and specialist services, such as the Improving Access to Psychological Therapies programme or Children and Adolescent Mental Health Service (CAMHS) Inpatient Services are usually provided by hospitals, mental health trusts or specialist delivery partners like Mind. The NHS collects real world data from people’s interactions with these services, in order to make evidence based decisions about healthcare, at both an individual and policy level. There are also a number of charitable organisations who offer mental health services to young people for free, particularly within schools, but also in an online format.

The UK invests around £125 million a year into mental health research, which is key to improving our understanding of mental health problems and creating effective treatments.\(^\text{43}\) About 9% of this spend contributes to depression focused research and just 4% looks at anxiety focused research. The research areas that receive the most funding (around 50% of total funding) tend to focus on underpinning research and aetiology. “Underpinning research aims to understand the structure and function of the brain and how it works in a healthy, functioning way – while aetiology looks into the factors contributing to the development of mental health conditions.”, The results of these projects can be used to conduct further research into treatments, interventions and the management of mental health conditions. The majority of this funding goes to academic institutions, such as universities, to support a variety of research studies. Common study designs for anxiety and depression research include longitudinal cohort studies, which observe the development of a group of people over a period of time; cross sectional studies, which look at a defined population at a single point in time; case-control studies, which examine multiple individuals with a shared trait or mental health condition and try to ascertain differences between their experiences; and systematic reviews, which provide a critical assessment of existing literature on a specific subject.

There are a number of commercial organisations that collect data about people which is relevant to mental health. This includes companies that provide sensors that can track biological changes in an individual; social media companies which capture information about people’s past and current thoughts and feelings, both publicly and through their online activity; and providers of mental health and wellbeing applications, which individuals can use to keep a record of how they are feeling about a particular topic or more generally.

All of these different types of data stewards could play an important role as potential data providers for the Wellcome mental health data challenge.

\(^\text{42}\) Full Fact. Mental health spending in the English NHS. https://fullfact.org/health/mental-health-spending-england/
Types of data

Symptoms
There are many different types of data that would be useful in better understanding the symptoms of anxiety and depression. Some examples of data sources for key types of symptoms include:

<table>
<thead>
<tr>
<th>Symptom type</th>
<th>Data source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physical: e.g. trouble sleeping</td>
<td>Sensor data (e.g. Fitbit)</td>
</tr>
<tr>
<td>Behavioural: e.g. lack of appetite</td>
<td>Self reporting / Observational data</td>
</tr>
<tr>
<td>Emotional: e.g. persistent low mood</td>
<td>Self reporting / Commercial apps / Patient records</td>
</tr>
<tr>
<td>Social: e.g. avoiding social activities</td>
<td>Social Media / Sensor data (Fitbit)</td>
</tr>
<tr>
<td>Cognitive: e.g. persistent negative thoughts</td>
<td>Patient records / Sensors (e.g. keystroke data) / Self reporting data</td>
</tr>
</tbody>
</table>

There are already numerous datasets that could help us to examine the symptoms of anxiety and depression. This includes data from the NHS, which can be accessed through NHS Digital’s mental health data hub, research data stewards (eg SAIL, Discover-NOW Hub, Adolescent Data Platform) as well as multiple alternative data sources such as charities, social media (e.g. Twitter), sensors (e.g. Fitbit or RADAR-CNS) and commercial mental health applications (eg Calm, Sleepio, Mind, Headspace).

Longitudinal studies such as ALSPAC, Neuroscience in Psychiatry Network and IMAGEN could help us to examine how young people’s mental health experiences change over time.

There is some research that has already been exploring how symptoms interact with each other. The challenge could build on this existing body of research on the interconnectedness of symptoms in the psychopathological network. Much of the data used in these examinations is openly accessible online through platforms like osf.io.

Core components of interventions
A core components focused challenge would seek to use data about interventions, from multiple sources. Some of the existing data (e.g. Discover-NOW Hub, IESO Digital Health, SHOUT / Mental Health Innovations, schools data in England, Wales and Scotland) is available for access by researchers, under different data access models. Other useful datasets exist (e.g. social media, sensor data from wearables such as Fitbit and commercial mental health applications like Calm, Sleepio, Mind and Headspace) but these datasets would be more difficult to access as part of the challenge.

Example data sources for mental health interventions:
Traditional clinical interventions

- Discover-NOW Hub has some of Europe's largest mental health service evaluation datasets through collaborations with NHS mental health trusts. They are currently looking at how to collect more real time data around mental health services and how to get more patient generated data that can support their analyses. They have already agreed in principle to share data as part of the challenge.
- NHS Digital has a number of datasets available in their mental health data hub about psychological therapies and mental health services. There could be potential to access GP data or specific data from the IAPT dataset.

Digital interventions

- IESO Digital Health has one of the most comprehensive mental health therapy datasets in the world, covering more than 250,000 digital therapy hours. They are interested in exploring a collaboration with us but are likely to only make their own data available if the data scientists are working on their premises and towards a specific challenge that they are already trying to address. They are keen to enable more effective interventions to be produced and have already seen recovery rates improve over time due to the insights they have drawn from their own datasets around what makes therapies work.
- SHOUT has thus far recorded data from over 250,000 conversations with young people who are in crisis. They would be willing to share much of this data with researchers as they want to get the best value from the data that they have.

Mental health and wellbeing interventions

- Different organisations in Wales (DECIPHer), Scotland (SHINE) and England (Place2Be) are collecting data about school mental health and wellbeing initiatives. Much of this data is available for researchers to access on a permission basis.

Limitations

Researchers are often unaware of similar research projects or datasets that exist outside of their sphere of influence. Much of the data that exists in the academic space is kept in silos, and therefore cannot be used by others to create new insights or make better decisions.

It can be very difficult for mental health researchers and data scientists to access data that they do not already steward. Some researchers suggested that it can take up to a year to access data from other research projects or the NHS.

Some data stewards are willing to share data with other researchers, but do not have the infrastructure in place to be able to enable safe access to data for people outside of their organisation.

Some organisations are reluctant to provide access to the data that they collect, or to enable that data to be linked to other data sources, because the data is about people and they are concerned about individuals retaining their anonymisation status.

A lot of commercial companies, in particular social media and sensor companies, are reluctant to share data that they collect as they see no value for themselves in sharing with external organisations.
Challenge structure

The challenge period will cover a period of 18 months, with a 9 month set up phase preceding the prize, and 9 month sustainability prize following the judging.

As previously mentioned, this challenge will need to be run in the UK, with a specific set of end user beneficiaries and community representatives which need to be engaged with throughout the design and delivery of the challenge prize. Participants should be recruited from across the country, and supported to travel to key events if necessary. This is intended to enable a peer network of local experts to play a core role in the challenge.

Seed funding would be needed in order to allow for people to form multi-disciplinary teams and take part in the competition. Details of this are available in the budget and in the descriptions below.

Training and mentoring would also be required, to align capacity and to ensure benefits would be lasting for the participants and the country as a whole. Specific training requirements will be established early in the challenge, depending on what participant needs are. Training might cover relevant areas from the ODI data skills framework44, such as ethical use of data or machine learning approaches.

Remit and scope of the challenge

Key question
How can we support the development of more effective interventions for addressing anxiety and depression in young people and adolescents (ages 14-24)?

Suitable proposals
Proposals should be submitted to make progress in one of the following areas:

- Better understanding the role that individual symptoms play in the development of other symptoms of anxiety and depression, which could help clinicians, researchers and individuals to improve existing, or develop new, mental health interventions which address those symptoms. By ‘symptoms’ we mean the physical, cognitive, emotional and behavioural patterns or manifestations of anxiety and depression.
- Better understanding the core components that make interventions for anxiety and depression work effectively, which could help clinicians, researchers and individuals create better interventions for young people who are at risk of developing, or have been diagnosed with, an anxiety or depressive disorder. By ‘core components’, we mean the active ingredients that make an intervention work, e.g. doing something on a daily basis

Any proposal submitted must aim to create tools or products that enable clinicians, other mental health care providers and individuals themselves to make better decisions about the

44 https://thecdi.org/article/open-data-skills-framework/
right treatment approaches for anxiety and depression in young people and adolescents. This can include tools that provide recommendations for interventions, approaches and techniques for young people and adolescents in general as well as tools to support decision-making for individual patients.

Proposals must consider the end goal of improved mental health outcomes through transforming treatment of depression and anxiety in young people in the UK.

Who can apply
We are looking for multidisciplinary teams who bring data science together with experience from the mental health space, in order to deliver innovative solutions to this complex issue.

Teams will need to be prepared to work in collaboration with mental health practitioners and young people in order to deliver suitable outcomes for these key beneficiaries.

The application process is open to those who have predefined ideas and projects, and to those who may have relevant skills but no project proposal.

Participant motivations
The motivation for participants to enter will be a combination of:

- opportunities for networking / building relationships
  - access to key government and healthcare sector stakeholders
  - access to other experts within and across disciplines
  - access to commercial partners and clients
  - access to beneficiaries
- personal and professional development
  - advancing the field of mental health research
  - advancing technological development
  - advancing individual capabilities
- personal and professional rewards
  - financial reward of winning the prize
  - reputational reward of winning the prize
  - potential to have their outputs implemented and create sustainable impact

The motivation for academic participants tends to be more driven by access to key government and healthcare stakeholders, improvement in and contributions to their field of study, and reputational awards.

The motivation for commercial participants tends to be more driven by access to new markets and clients, positive press around service provision and financial rewards.

The motivation for governmental stakeholders will be to outsource the development of innovations that can help them deliver more efficient services, and contribute towards SDG goals.
Challenge stages, funding and support

The estimated time allocations for the mental health data challenge would include:
- Up to **9 months** for a set up phase
- A minimum of **18 months** for the challenge to run, from recruitment of participants through to the awarding of the challenge prizes
- Up to **an additional 9 months** for a sustainability period

For a detailed view of estimated timelines, please refer to the mental health data challenge timeline in Appendix 10.

1. **Set up phase**
   The organisation that is selected to run this challenge programme would likely need up to nine months in order to set up the challenge programme. This might include, but is not limited to, the following activities:

   1. Convening beneficiaries and stakeholders across the communities in question, local, state and national government, private sector and third sector in order to verify their needs with regard to the mental health focus areas and confirm their engagement with the prize.
   2. Convening community representatives to understand how municipal needs relate to community needs.
   3. Establish agreements with local delivery partners, and establish the detailed delivery plan.
   4. Establish agreements with data providers. Depending on the available data in the chosen location, some ground truthing / community data may need to be collected prior to launch.
   5. All data will need to be vetted and cleaned thoroughly prior to launch in order to verify quality and limitations. This will take dedicated resources and be the most time consuming part. The aim is to ensure that participants can use the data immediately at the point they enter the insights phase.
At the end of the set up phase, the expression of interest phase would be initiated, either immediately or after a set amount of time.

2. **Participant recruitment/ Expression of interest**
   A website or landing platform will need to be created with all the required information on topic and processes. This will then be publicised through some advertising and media channels. It is likely that the bulk of the participants will be recruited through established networks such as the Transdisciplinary Research for the Improvement of Youth Mental Public Health (TRIUMPH), Emerging Minds: Action for Child Mental Health, the Alan Turing Institute and the MQ Mental Health Data Science Group.

   Participants will be invited to complete an expression of interest application over a four to eight week period. EOI’s can be open to those who have predefined ideas and projects, and those who may have relevant skills but no project proposal. Participants will be able to communicate with each other during this period, and will be encouraged to make connections and discuss ideas to help with team formation.

3. **Launch lab**
   The challenge prize will begin with a ‘lab’ or multi-day event (possibly virtual given the Covid-19 crisis). All those who have completed an EOI will be invited, along with beneficiaries, data providers and delivery partners. Participants will have their expenses paid for travel and accommodation.

   The first part of this event will focus on explaining why this challenge topic is so important, telling the stories of why and how this problem needs to be addressed, and the processes involved in participating in the challenge prize. The event will also cover why there are two challenge prize streams and what the differences are between the two. This can be recorded to be sent to people who cannot attend. Press will be invited to cover the ‘launch’ and meet some of the participants and key stakeholders.

   Following this introductory session, participants will have the opportunity to share their project ideas, recruit new team members, and engage with data providers and stakeholders.

   The aim is to support the development of multidisciplinary teams, bring them into contact with the data providers and beneficiaries, and help them decide which challenge prize stream might be of interest for them to submit a proposal for. There will be a clear process for how to team up and formally enter as a team.

   The teams will be given four to seven weeks following the launch to confirm their teams and submit their area of interest / initial proposal in a formal team application.

4. **Team applications**
   The teams will be given 4-8 weeks following the launch lab to refine their proposal ideas, confirm their teams and submit their area of interest/ initial proposal in a formal team application. This will then be assessed based on criteria including:
   - Multidisciplinary team with relevant expertise
   - Area of interest in line with challenge scope
   - Proposed engagement with end beneficiaries
- Scale of ambition

During this period, participants will have access to a challenge team application template to complete, have access to supporting information on the topic, and will be able to troubleshoot with the organisation running the challenge for process queries.

Teams may submit their proposal at any time during this period and begin the next phase, though there will not be access to funding before 4 weeks. The application window will remain open until 8 weeks after the launch lab.

The application would also need to clearly state whether the project is for the core components challenge prize or the symptoms challenge prize.

5. Model development phase

Successful applicants will enter into phase one of the challenge, otherwise known as the model development phase. Both the core components challenge prize stream and the symptoms challenge prize streams would run alongside each other from this point onwards. This phase will be between six to nine months and focus on the development of data insights. The teams will be encouraged to focus on engaging with the available data and modeling it to develop insights. In this phase, teams would receive some seed funding (£25k for 10 groups in each stream), data access support to help them identify the data they need for their project, as well as training on relevant areas from the data skills framework, such as ethical use of data or machine learning approaches.

The exact details of what each team plans to achieve by the end of the phase should be outlined in the individual project plan, and will therefore be pre-agreed at the start of the phase with each team. In this phase we would expect the members of the teams with data science and mental health expertise to be leading the work.

Through the development of models and insights, the teams will be asked to put together plans for developing practical applications and tools for end-users drawing on the data science methods and insights they have developed. These tools and applications will need to address specific needs of beneficiary groups such as, but not limited to:

- Clinicians or other mental health care providers who need to make better decisions about the right treatment approaches for anxiety and depression in young people and adolescents
- Young people with lived experience of anxiety and / or depression, who need to make better decisions about their approach to their own mental health and wellbeing

As such, participants would be expected to engage with one or both of these stakeholder groups. Introductions to the trusted intermediaries, who can facilitate access to these stakeholders groups would be provided upon request.

The model development phase will culminate in a demo day in which all of the teams will demo the models and insights they have developed using the data, and their pitches for tool development. These will be judged by a panel of key stakeholders and experts in the field against a set publicly accessible scoring criteria. For example, applications might be assessed on some of the following areas with a focus on
evaluating the innovation of the data modeling done so far, the feasibility of the proposal and the potential impact:

- Understanding of the problem
- The proposed approach
- Team resources
- Proposed outcomes
- Potential impacts
- Project management documentation
- Risk assessment
- Cost plan
- Ethical assessment

The best teams will be selected to go through to the development stage to build their proposed tools / applications. The number of successful applicants who move from the insights phase to the development phase would likely be pre-agreed at this stage, in order to control costs in the discovery phase.

6. **Tool development phase**

This phase of the challenge would run for six to nine months. Both the core components challenge prize stream and the symptoms challenge prize streams would continue to run alongside each other. At this point, the projects that have progressed will be offered the chance to create their tools or applications using an additional amount of seed funding (£100k for up to 5 teams in each stream).

In this phase we would expect the teams to have a greater focus on digital development, working with and drawing on the expertise of the team’s scientific experts and stakeholders to inform their work. These teams will receive training and mentorship suitable to their specific topic to enable their ideas to develop and flourish. This will include support to plan implementation with stakeholders, ethical assessments and business model development.

Where possible, the challenge teams will get a chance to implement their full or partial solutions in an environment with real stakeholders. The primary stakeholders here will either be mental health practitioners, or young people with lived experience of anxiety and / or depression. This will allow for feedback, adjustments and further development before the final judging. This will need to have been agreed in principle with key stakeholders during the implementation phase, and then coordinated by the delivery team following the model development period. For both of these key stakeholders, it is important that participants work with them to make sure that the tools that are created are both beneficial to them and work in practice.

7. **Judging**

The development phases will culminate in a public demo day attended by as many media representatives and senior stakeholders attending as possible. Each of the teams present their work. The event will have a strong focus on storytelling - what the nature and scale of the risks around urban public health are, and how we have used data science to address them.

A set of judges who are well known in the field, such as Paul Farmer (CEO of Mind), Mark Rowland (CEO of Mental Health Foundation), Matt Hancock (Secretary of State for Health and Social Care), Simon Stevens (CEO of NHS England) will review the submissions against publicly available criteria. Criteria are likely to include:
● Innovation
  ○ Ethical data management
  ○ Utilises multiple data sources
● Impact
  ○ Scale of potential impact on urban health in the region
  ○ Building trust with the public on the use of data for health innovation
● Capacity
  ○ Clear input from multiple disciplines
● Beneficiaries
  ○ Clear link to end users
  ○ Scalable solution which has the potential to be used in other countries
● Sustainability
  ○ A clear, time bound implementation plan for the solution to be adopted by end users
  ○ Whether the solution has the potential to continue beyond the prize

The most successful submission in each stream (core components and symptoms) will be awarded a prize of £200k. Three runner up teams in each stream shall receive £100k. This money will be explicitly for the implementation and scaling of the tool they have developed. The rationale for this model is that it allows multiple tools and outputs to go on to be scaled and supported, delivering sustainable impact in the field. This has been benchmarked against other health technology innovation prizes.

8. **Sustainability/ Uptake period**

It is vital that there is a sustainability period following the award of the prize in order to support the winning project and the other projects to be adopted by the key stakeholders and used to address the key issues addressed regarding mental health. The winning teams can use the prize money to implement and scale up the innovations they have developed.

In addition, the partners involved in the delivery of the prize will need to remain available for the teams over a period of 6 month in order to:

● Assist the future development of winning solutions by making introductions to would-be users, investors, commissioners and others.
● Establish a schedule for checking progress with winners and interesting other participants.
● Seek feedback from competitors, judges and partners and evaluate the effectiveness of the programme.
● Return to the prize winners a year afterwards to develop a case study of where they have got to in their development and the impact they have created.

9. **Evaluation**

At the end of the judging and sustainability phase there shall be an independent evaluation made of the prize as a whole drawing on reporting material and interviews with participants and stakeholders. The interviews will also be used to develop case studies of each output and prize in order to share learnings and successes for others interested in the field of health data innovation and remote sensing for public health.

45 See Appendix 9
Challenge prize data ecosystem

Mental health challenge data ecosystem map. A high res version is available in Appendix 11, and an interactive version is available [here].

This data ecosystem map displays: the potential data flows to challenge participants addressing anxiety and depression in young people; the potential value flows to the healthcare and research data providers from the challenge participants; and the value flows to and from organisations who aren’t directly sharing data, such as young people networks and researchers external to the challenge.

The map provides an example of some of the data sources which a team of participants would be able to draw on in order to create solutions to address one of the challenge focus areas. There are many potential data sources which have been included in the data landscape summary, but may not be included in the map, in order to make sure that the map is legible.
Risks

There are three major risks associated with this challenge.

Firstly, impact can only be created from this challenge if participants are able to get access to good quality data about the symptoms of anxiety and depression, and mental health interventions. Mental health data has traditionally been very difficult for researchers to access, and the COVID-19 situation may present issues with being able to access certain datasets, if data providers are unable to provide remote access to the data. We aim to mitigate this risk by creating formal agreements with data providers in the set up phase of the challenge.

Another potential risk to the challenge could be that participants are unable to meaningfully link datasets from different, which could come as a result of incapable data infrastructure, or a reticence from data providers because of the nature of the data, to allow data linkages. We aim to mitigate this by having a clear understanding of what each data provider is willing and capable of allowing participants to do in terms of linking datasets together, and giving participants and data providers the space to have conversations about the subject.

Additionally, one risk with working with mental health data, and data about young people, is that participants aren’t handling or using data in an ethical way. We aim to mitigate this by making sure that young people are engaged regularly throughout the challenge prize, in order to make sure they are comfortable with the use of data and outputs that are being created. We will also provide teams with support and training around ethical use of data.

For a detailed breakdown of the potential risks, please see the risk register for this challenge.\(^\text{46}\)

The risk register is not exhaustive and factors may change over time, particularly with the changing legal, policy and regulatory environment in light of the current COVID-19 situation, and in the longer term, Brexit.

Capacity and stakeholder landscape

The UK has been selected as the focus for this health data challenge, because of its strong data infrastructure, availability of data, mental health expertise, regulatory frameworks and the connectedness of the UK’s public healthcare system, all of which provide a strong base with which to build the foundations of the challenge on.

The challenge focus is one that could be applicable to other geographies, but it would be remiss to try and address these challenges in other countries, while the UK has yet to address them. There are also very few countries that have the wealth of domain experience, health data and data science capacity. This challenge could be scaled to other geographies after completion, but attention would need to be given to the state of data infrastructure, availability of data, as well as mental health expertise and data science capacity available.

Below is information on mental health stakeholders in the UK. In terms of recruitment, there are a number of data science and public health networks across the UK. ODI’s recommendation is to focus on building connections with these domestic networks, rather than supporting the wider international community.

\(^{46}\) Appendix 3
Mental health researchers

There are numerous academic researchers in the UK who have mental health research expertise and most of the mental health research funding goes to academic institutions. However, mental health research receives a comparatively small amount of funding (£125 million a year) compared to other major health research areas, such as cancer research (£404 million a year). This may be, in part, due to how reliant the mental health space is on government grant funding, as 67% of mental health research is funded by the UK Government, with around 30% being provided by other funders, such as philanthropic organisations, and just under 3% coming from public donations collected by charitable organisations.

Mental health research is vital in enabling us to get a better understanding of mental health disorders, which can help to make interventions more effective and ultimately make people’s lives better, therefore it is important that funders continue to support this work as much as possible. The Wellcome mental health data challenge could make a significant contribution to supporting research advancements in our understanding of anxiety and depression symptoms and interventions, and making sure that researchers from mental health backgrounds are partners in the challenge will be key.

There are a large number of strong mental health networks in the UK, such as Transdisciplinary Research for the Improvement of Youth Mental Public Health (TRIUMPH), Emerging Minds: Action for Child Mental Health, SMARTEN: Student Mental Health Research Network and The Nurture Network: Promoting Young People’s Mental Health in a Digital World. These networks will be key in both recruitment of participants for the challenge prize, and engaging with potential adopters of the various tools and applications, such as clinicians and other mental health practitioners.

Stakeholders that we have engaged with in this area include:

- Discover-NOW Hub
- Francis Crick Institute
- Imperial College London
- Imperial Centre for Health Policy
- Professor Ann John, Swansea University
- Professor Tamsin Ford, Cambridge University
- Professor Simon Murphy, Cardiff University
- Professor Andrew McIntosh, Edinburgh University
- Professor Glyn Lewis, University College London
- Innovate UK
- UK Research and Innovation
- Medicines Discovery Catapult

---

Beneficiaries

This health data challenge will require targeted engagement with potential beneficiaries. This is integral in making sure that the outputs that are produced are useful in helping them to make better decisions about the anxiety and depression interventions.

There are two main groups of beneficiaries:

**Mental health practitioners who are responsible for supporting young people with mental health problems.**

Mental health practitioners develop, prescribe and implement mental health interventions. Practitioners can either be from a primary care background, such as your GP, or a specialist mental health background, such as a clinician or therapist. These individuals need to know what symptoms an individual might have, how these symptoms can be addressed through the application of a specific intervention and what this intervention needs to incorporate in order to be suitable and effective for that individual. Networks with connections to mental health practitioners might include:

- [Child Outcomes Research Consortium](#)
- [Mental Health Network](#)
- [Clinical Commissioning Groups (CCGs)](#)

**Young people (ages 14-24) with lived experience of anxiety and / or depression, who need support to better manage their symptoms.**

It is not always necessary for a young person to seek professional support and some intervention or prevention techniques can be self led. For example, some individuals might find that milder symptoms can be managed effectively through exercise, meditation or social engagement. However, young people need the tools to be able to help them decide what the best intervention approach might be for them as an individual. It is critical that these people are engaged through a trusted intermediary, so that they can share their experiences and needs, while also feeling safe and comfortable to do so. Networks with connections to young people might include:

- MQ Mental Health, via [Participate](#)
- Place2Be
- [Transdisciplinary Research for the Improvement of Youth Mental Public Health (TRIUMPH)](#)
- [Young Minds](#)

**Third sector influence**

There are a number of charities and volunteer organisations in this space who are supporting research and convening people around problems. Some of these organisations, like Place2Be, use their funding to provide practical support to those in need. Others, like MQ Mental Health, provide funding to researchers in order to advance the research agenda in the mental health space. Many third sector mental health organisations also play a role as a convener, by bringing young people with lived experience of mental health issues to researchers and / or practical support.
These organisations are likely to play a key role in the mental health data challenge, both from a recruitment perspective and for their ability to connect challenge participants with young people. It is therefore important that some funding is allocated to partnerships with third sector organisations.

Stakeholders that we have engaged with in this area include:
- Place2Be
- MQ Mental Health
- Mind
- SHOUT / Mental Health Innovations UK

Industry influence

The adoption of digital technology is becoming increasingly common and many of the major developments in technology are being driven by the private sector. These companies are important to the mental health space because of the data that they collect about people.

Sector agnostic technology companies, such as large social media organisations, have incredibly useful datasets about mental health, because of the way that social media users report about themselves in real time. Sensor providers in the health space are capturing real time data about an individual’s physical state. Wellbeing applications afford people the opportunity to talk about how they are feeling, and monitor how that changes on a daily basis.

It can be very difficult to make the case to these companies that they should be as open as possible with the data that they collect, as there could be an argument made that there is no commercial value to be had in sharing this data, or that they do not have permission from users to share data about them. From our conversations with experts in this area, the main justification for large technology companies to provide access to data is to help solve a global problem - as is evident with the COVID-19 pandemic - or to strengthen their position in a particular market - for example, positioning themselves as a technology provider to a specific healthcare provider like the NHS.

Encouraging these large commercial organisations to increase access to data, in a way that still retains the trust of their customers, would enable researchers with mental health expertise to access rich data sources that would benefit research. Wellcome, and other influential funders in the mental health space, could use their credibility and profile to negotiate access to some of these datasets. Stakeholders that we have engaged with in this area include:
- IESO Digital Health

Wider challenge landscape

Through our research, we have not identified a large amount of data science initiatives in the UK mental health space. Below we have listed a few of the existing initiatives:

- **The Alan Turing Institute** - [Data science for mental health](#)

The Alan Turing Institute has created a network of people who are interested in collaborating on data science projects within the mental health space. The collaborations have so far looked at
areas such as cognitive science, social psychology and developmental psychology, although the latter area is quite a new focus.

**MQ Mental Health Data Science Group**
MQ Mental Health has convened a group of mental health experts who are focused on championing data science in mental health, mapping the data science landscape, and trying to improve access to data for mental health researchers. The group has published a paper: [Data science for mental health: a UK perspective on a global challenge](#).

**Mental Health Data Science Scotland**
Through this project, a number of academic researchers in Scotland are examining big datasets in order to try and better understand the mechanisms, diagnosis and treatment of common mental health conditions.

**Legal, policy and regulatory environment**
The ODI subcontracted Future Care Capital to undertake a review of the policy, regulatory and legal environment which encompasses mental health, and health more broadly, in the UK.

The review looks at six different aspects:

1. Data Protection: legal and regulatory framework
2. Health Data: legal and regulatory framework
3. Health Data Controllers in England: high-level landscape review
4. Local Health Data Controllers in England: state institutions
5. Intellectual Property: legal and regulatory framework for data-driven health

The health data landscape in the UK is very complex, with thousands of health data stewards ranging from government agencies to public-private research initiatives to third sector healthcare providers, and more. All of these organisations will have to comply with European and national regulations for health data. According to the review, there are many organisations involved in monitoring compliance with regards to the collection and use of health data in the UK, including NHS Digital, National Data Guardian and National Institute for Health Care Excellence. These organisations are not the same as those who are responsible for enforcing data protection laws (specific or non-specific), which includes the Information Commissioner’s Office and the Health Research Authority. In order for the mental health data challenge to avoid creating negative impacts, the organisation responsible for running the challenge must establish relationships with key regulators during the set up phase of the challenge.

Depending on a number of factors, such as the sector the organisations operates in, or the location or nature of the communities that they serve, each health data steward will also have other rules and guidelines to follow, including their own internal data governance requirements. Due to the potential complexities in how data is accessed from each individual organisation, it is integral that agreements with data providers are put in place before the challenge is launched, so that potential participants can access the data from the point of launch.
An important consideration for the challenge is understanding who the key beneficiaries are and what their needs might be. ‘Young people’ could be considered a fairly broad audience, and there are bound to be variations in how different young people interact with the UK healthcare system. For example, not all UK inhabitants have the same access to healthcare services. Many of those with refugee status are not entitled to free secondary care for mental health, which could mean that they are not represented by the data which is held by these institutions. This could mean that the outcomes from the data challenge might not be beneficial to this particular community. Giving potential participants a view of the available data before the challenge is launched is important, in order for them to be able to understand the opportunities and limitations within the data. Regular public engagement with a diverse representation of young people throughout the challenge will also play an important role, in making sure that the needs of end beneficiaries are being considered in both the way that the challenge is structured, and the proposed outputs from each participant group.

Public engagement is not just a tool for understanding the needs of beneficiaries. It is also important to understand how potential beneficiaries feel about the use of data about them. Using health data in ways that are opaque or perhaps harmful to individuals can be detrimental to both the success of the initiative and to the individuals themselves. For example, Project Nightingale, a collaboration between healthcare provider Ascension and Google, recently came under fire for when a whistleblower suggested that the personal health records of over 50 million Americans had been shared before the data had been de-identified,49 meaning that a number of Google employees had access to a number of personal medical histories without the individual’s consent. This error has led to subsequent legal proceedings against both companies. In contrast, other initiatives have demonstrated that connecting with the public about how data can be used to support their health needs can lead to a shared understanding about, and support for, the impact you are trying to achieve. The Connected Health Cities initiative is a great example of how earning citizens trust by involving them in the design process can help to save lives, build capacity and make public healthcare more cost effective.50

In summary, despite some of the complexities associated with the mental health space in the UK, the review suggests that a data challenge would be feasible if the following areas are considered before launching the challenge:

---

These suggestions align with our proposed challenge approach. The organisation who will run the challenge phase will need to work with data stewards / controllers to make sure that access to data is provided from the moment that the challenges begin, which is part of the justification for the proposal of a period of up to nine months to set up the challenge. Regulators and governing bodies will also need to be engaged during this time, to make sure that the proposed use of data is lawful and ethical. The challenge will encourage innovative thinking by enabling the formation of multidisciplinary teams, with different experiences and skill sets. Introductions will be facilitated through the launch lab and team formations phase, where many of these different actors will be invited to come together to find out more about the challenge. Participants will also be required to submit proposals, so that project ideas can be vetted against a predefined criteria. This will make sure that participant proposals are aiming to create a positive impact in the mental health space, and that the use of data will be to the benefit of young people with anxiety and depression.

The full review of the mental health policy, regulatory and legal environment in the UK is available in Appendix 15.

## Data Inventory

<table>
<thead>
<tr>
<th>Source</th>
<th>Description</th>
<th>Where does this data sit on The Data Spectrum?</th>
<th>Focus area</th>
</tr>
</thead>
</table>
| Discover-NOW Hub | Service evaluations from mental health trusts and on mental health conditions and symptoms from GP records  
One of Europe’s largest linked dataset at patient level including primary, secondary, tertiary and social care data. | Shared - Named access  
Have agreed to have further conversations about being a data provider - Memorandum of Understanding [here](#) | Core components  
Symptoms |
| Place2Be        | 35,000 conversations with a few data points  
A smaller number of detailed, high dimensional datasets.                                                                                                                                                   | Shared - Named access | Core components  
Symptoms |
| SHOUT           | Crisis text line service. Has collected data from around 250,000 conversations with young people about their mental health, including depression and anxiety                                                                                     | Shared - Named access | Core components  
Symptoms |
| IESO Digital Health | Provides 1:1 cognitive behavioural therapy online. Collecting data from around 50,000 conversations. Also tracking recovery rates                                                                                   | Shared - Named access | Core components  
Symptoms |
<table>
<thead>
<tr>
<th>Source</th>
<th>Description</th>
<th>Access</th>
<th>Core components</th>
<th>Symptoms</th>
</tr>
</thead>
<tbody>
<tr>
<td>DECHIPHer</td>
<td>Collecting data about school mental health and wellbeing in Wales</td>
<td>Shared - Group-based access</td>
<td>Core components</td>
<td>Symptoms</td>
</tr>
<tr>
<td><strong>Secure Anonymised Information Linkage (SAIL) Databank</strong></td>
<td>A collaboration between researchers in Wales to anonymise and link routinely collected data from the NHS and other sources in a safe and secure way, for the purposes of supporting health and social care research. Some datasets are publicly available, whilst others are permission based</td>
<td>Shared - Public access</td>
<td>Symptoms</td>
<td></td>
</tr>
<tr>
<td><strong>Adolescent Mental Health Data Platform</strong></td>
<td>A platform that holds a number of mental health datasets about primary and secondary care, as well as demographic data. Some datasets are publicly available, whilst others are permission based</td>
<td>Shared - Public access</td>
<td>Symptoms</td>
<td></td>
</tr>
<tr>
<td>NHS Digital</td>
<td>Has a number of datasets available in their mental health data hub about psychological therapies and mental health services.</td>
<td>Shared - Public access</td>
<td>Core components</td>
<td></td>
</tr>
<tr>
<td><strong>NHS Digital</strong></td>
<td>Collects data about the Adult Improving Access to Psychological Therapies (IAPT) programme, which will include information about young adults (aged 18-24). All IAPT services collect a Minimum Data Set (MDS) about individual patient progress and outcomes of treatment.</td>
<td>Shared - Named access</td>
<td>Symptoms</td>
<td>Core components</td>
</tr>
<tr>
<td><strong>Schools Health and Wellbeing Improvement Network (SHiNE)</strong></td>
<td>Collecting data about school mental health and wellbeing in Scotland</td>
<td>TBC</td>
<td>Core components</td>
<td>Symptoms</td>
</tr>
<tr>
<td>Remote Assessment of Disease and Relapse – Central Nervous System (RADAR-CNS)</td>
<td>A collaborative research programme exploring the potential of wearable devices to help prevent and treat depression</td>
<td>Shared - Named access</td>
<td>Symptoms</td>
<td></td>
</tr>
<tr>
<td><strong>Open Science Framework</strong> (by the Centre for Open Science)</td>
<td>A free platform that allows health researchers to look for psychology data and projects that other researchers have made available</td>
<td>Open</td>
<td>Core components</td>
<td>Symptoms</td>
</tr>
<tr>
<td>National Institute of</td>
<td>De-identified human subjects data</td>
<td>Shared - Group-based</td>
<td>Symptoms</td>
<td></td>
</tr>
<tr>
<td>Mental Health Data Archive</td>
<td>from mental health research is available to qualified researchers. Summary data is publicly available.</td>
<td>access</td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------------------------</td>
<td>-------------------------------------------------------------------------------------------------</td>
<td>-------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PsyArXiv</td>
<td>A free preprint service for the psychological sciences</td>
<td>Open</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clinical Record Interactive Search (CRIS)</td>
<td>Provides authorised researchers with regulated, secure access to anonymised information extracted from South London and Maudsley NHS Foundation Trust electronic clinical records system.</td>
<td>Shared - Group-based access</td>
<td>Core components</td>
<td>Symptoms</td>
</tr>
<tr>
<td>GLAD study</td>
<td>Longitudinal study looking at depression and anxiety risk factors and genetics links.</td>
<td></td>
<td>Symptoms</td>
<td></td>
</tr>
<tr>
<td>Avon Longitudinal Study of Parents and Children</td>
<td>Birth cohort study looking at the environmental and genetic factors that affect a person’s health and development.</td>
<td></td>
<td>Symptoms</td>
<td></td>
</tr>
<tr>
<td>Born in Bradford</td>
<td>Longitudinal study tracking children’s health and wellbeing from birth, through adolescence and into adulthood</td>
<td></td>
<td>Symptoms</td>
<td></td>
</tr>
<tr>
<td>Determinants of Adolescent Social Well-being and Health (DASH)</td>
<td>DASH looks at how social, environmental and biological factors affect the health and well-being of young people from different social and ethnic backgrounds in London.</td>
<td></td>
<td>Symptoms</td>
<td></td>
</tr>
<tr>
<td>Growing up in Scotland</td>
<td>A longitudinal research study, tracking the lives of thousands of children &amp; their families from the early years, through childhood and beyond.</td>
<td></td>
<td>Symptoms</td>
<td></td>
</tr>
<tr>
<td>Millenium Cohort Study</td>
<td>A longitudinal study following the lives of around 19,000 young people born across England, Scotland, Wales and Northern Ireland in 2000-02. The MCS provides multiple measures of the cohort members’ physical, socio-emotional, cognitive and behavioural development over time, as well as detailed information on their daily life, behaviour and experiences. Alongside this, rich information on economic circumstances, parenting, relationships and family life is available from both resident parents. National Pupil Database (NPD) records have also been linked to the MCS data, including GCSE exam results.</td>
<td></td>
<td>Symptoms</td>
<td></td>
</tr>
<tr>
<td>NeuroScience in Psychiatry Network</td>
<td>A longitudinal study researching how the adolescent mind and brain develops into early adulthood.</td>
<td>Symptoms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------------------------------</td>
<td>------------------------------------------------------------------------------------------------</td>
<td>----------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IMAGEN</td>
<td>A European research project examining how biological, psychological, and environmental factors during adolescence may influence brain development and mental health.</td>
<td>Shared - Group based access</td>
<td>Symptoms</td>
<td></td>
</tr>
</tbody>
</table>
Snakebite health data challenge: Detailed Scoping

Disclaimer

The content of this challenge topic aligns closely with the strategic objective and direction of Wellcome’s snakebite priority area. However, due to the likely impact of COVID-19, it is advised that the detail of this proposal is revisited in late 2020.

Summary Recommendation

Key challenge question

How can we use data science to reduce deaths from snakebite in India?

Why?

Snakebite kills more than 120,000 people and leaves 400,000 people with life changing disabilities every year globally, an equivalent global health impact to prostate or cervical cancer.\(^{51}\) With over 46,000 of these deaths occurring yearly in India, and three times that number left permanently disabled, India ranks highest in the world for snakebite burden of disease.\(^{52}\)

Deaths and disabilities from snakebite could be reduced through better access to data on the epidemiology of snakebite, leading to better access to proper treatment such as antivenoms. A study conducted in 2013 in Tamil Nadu notes that ‘Information about snakebite incidence is [also] lacking: there is insufficient epidemiological data, particularly in the rural areas where snakebites are most common.’\(^{53}\) In interviews, experienced field researchers have described snakebite as more than just a health and humanitarian crisis, but also a data crisis.

Addressing it requires an interdisciplinary approach, including more openness and sharing of data. Existing snakebite data alone is not enough to understand the full scope of this problem and to power its solutions. By combining data from outside of traditional research such as climate, ecology and urbanisation, and using data from a region that is perceived to have one of the strongest snakebite data infrastructure in the country, reliable models for mapping and predicting snakebite and envenoming can be created.

---

What?

The field of snakebite management holds a huge amount of potential for the use of data science, and the building of public trust, identifying correlations and risk factors through the analysis of medical, environmental and behavioural data. A challenge prize with a package of seed funding, training and sustainability support worth over £1.25million would be one way to encourage multidisciplinary teams to collaborate to use data science to provide the insights and tools decision-makers need to ensure their plans and responses are appropriate for the level of risk posed by snakebite in a geography.

A Health Data Challenge Prize run in Tamil Nadu could galvanise the research and data science community to create impactful solutions from existing yet disparate data sources. It would likely achieve the greatest impact if launched in early 2021 in order to build on and amplify the activities of international bodies such as the WHO, and research institutions such as the University of Geneva, while allowing focus in global health to move on from the COVID-19 pandemic.

The models created in Tamil Nadu could then be adapted to other regional contexts with more limited snakebite data, such as nearby Kerala, or the more far-flung northeastern states. These models could enable better prediction for snakebite instances, envenoming, mortality and morbidity and the identification of areas/populations at highest risk. This could allow for better preparedness through the optimisation of healthcare resources, such as antivenom supplies, and lead to lowered risks of mortality and morbidity in Tamil Nadu.

Where?

The state of Tamil Nadu, with both high snakebite prevalence and above average healthcare and data infrastructure represents an opportunity for a data science-focussed solution. Using existing data sources to identify and refine patterns and trends in snakebite incidence and prevalence in order to support the local healthcare system to predict and respond to snakebite envenoming, could help reduce the mortality and morbidity of this neglected tropical disease.

Who?

<table>
<thead>
<tr>
<th>Creators (Participants)</th>
<th>● Multidisciplinary teams with expertise in data science, public health &amp; biology (data scientists, clinicians, policy makers, biologists, virologists, epidemiologists), environmental sciences &amp; ecology (ecologists, environmentalists, herpetologists) could identify new insights and develop new tools based on data combined for the first time.</th>
</tr>
</thead>
</table>
| Decision makers        | ● Public health officials in charge of antivenom management could use the tools to optimise distribution  
● NGO leaders focused on prevention activities could use the tools to plan their public awareness and education campaigns  
● Health policy-makers could use the tools to inform long-term planning and resource distribution based on the potential impacts of climate |
<table>
<thead>
<tr>
<th>Beneficiaries</th>
</tr>
</thead>
<tbody>
<tr>
<td>● Victims of snakebite will have better access to medical services such as antivenom and ancillary treatment</td>
</tr>
<tr>
<td>● The at-risk populace will have better advice on preventing snakebite incidence in their communities</td>
</tr>
</tbody>
</table>

**Background Information**

Over the past few years, venomous snakebite has started to receive more attention as a significant global health problem. In 2017, the World Health Organization (WHO) categorised it as a neglected tropical disease (NTD) and committed to halving the number of deaths and cases of disability due to snakebite envenoming by 2030, with the pilot phase running from 2019-2022.  

54 Given this new classification snakebite is now linked to United Nations Sustainable Development Goal (UN SDG) 3.3 “By 2030, end the epidemics of AIDS, tuberculosis, malaria and neglected tropical diseases and combat hepatitis, water-borne diseases and other communicable diseases.”  

55 In addition, the Wellcome Trust made snakebite a priority area with the ambitions to modernise snakebite treatment production, develop the next generation of treatments, and build and sustain snakebite as a global health priority.  

56 Their priority region is Sub-Saharan Africa, with support in other regions such as South Asia aimed at identifying opportunities for ideas and technologies transfer to Africa.

In India, the prevalence of snake envenoming varies by state, with the northeastern states of Uttar Pradesh and Bihar bearing some of the worst of the disease, while official government statistics tend to understate the regional mortality by over 99% in some instances.  

57 The estimated number of death from snakebites in Tamil Nadu ranges from 3,000 per year in one study, 58 to upwards of 10,000 with 113,000 total snakebites based on community surveys in another.  

59 Research has shown four major factors that contribute to both a high prevalence of snakebite, and concurrent underreporting: a suitable habitat for snakes; no antivenom; poor quality health care; and inaccessible health care.  

60 These factors therefore tend to converge in rural and impoverished areas, where data collection is minimal, and where analysis of non-traditional datasets may be required to paint a better picture of the scope and severity of the issue.

---


56 Ibid


Researchers from multiple institutes around the world have been studying the snakebite issue in India, collecting a variety of data from hospital records and community surveys to labelled photographs of snakes themselves. At the University of Geneva for example, have been working on a medical decision-support tool for snake identification based machine learning algorithms fed by citizen-generated data.61 However the scope and variability of India’s geography, climate, and health infrastructure make providing national-level solutions difficult. In the southeastern state of Tamil Nadu, snakebite incidence is high, but unlike many other states, the quality of healthcare systems and snakebite research is high too, with sufficient data infrastructure for a challenge prize to demonstrate the potential of data innovation in snakebite management.

The snakebite health data challenge will be able to build off of the work and momentum of other organisations and activities seeking to reduce snakebite mortality and morbidity. International bodies like the WHO are leading the global drive against snakebite, while regional initiatives across South Asia and further abroad have helped paved the way for data-driven solutions. Challenge participants can look to these initiatives for support in terms of advice, data and inspiration.

Theory of change

A full theory of change is available as in Appendix 1. The underlying hypothesis for the challenge is:

- **If we:** run a challenge for multi-disciplinary teams on human, snake and environment interaction using a variety of both traditional and non-traditional datasets (such as environment, ecology and demographics),
- **Then:** data scientists should be able to use insights on the level of risk and potential burden of snakebite at a local level to develop products and tools,
- **Which:** could support healthcare providers (e.g: healthcare facilities/ hospitals/ department of health) and other actors involved in snakebite management to better plan and respond to snakebite instances, e.g. with better antivenom distribution and management as well as better distribution of other resources necessary to treat snakebites
- **Ultimately leading to:** a reduction of snakebite mortality and morbidity in India

Likely types of benefit

The challenge prize primarily aims at developing solutions for planning health services better by mapping areas and populations at highest risk of snakebite and envenoming and therefore better predicting the need for snakebite prevention and treatment. With better prediction of the incidence of snakebites, risk of envenoming and risk of mortality, the distribution and

---

management of antivenom and other resources needed to treat snakebite can be optimised.

Preventive interventions, such as community outreach programmes to educate people about measures they can take to lower the risk of snakebite, can also be better targeted based on more accurate local estimates of the risk of snakebite. The solutions to the challenge prize might include tools that draw insights from snakebite and envenoming risk maps/maps of unmet need of antivenom treatment or similar products that can inform the planning of healthcare service allocation or prevention efforts.

Through such challenge, teams would likely address the following specific issues:

• planning healthcare services better for more efficient response to snakebite incidence
• better targeting preventive care, e.g. public awareness campaigns that help individuals take protective measures to lower the risk of getting bitten by a snake, by the government and other stakeholders (e.g. nonprofits)

The type of decisions that could be made more effectively as a result of the tools, products and services developed include, for instance:

• Public health officials in charge of preparing an annual request for antivenom in a healthcare facility being able to order the right quantity of antivenom
• NGO leaders focused on prevention activities, e.g. community outreach programmes to educate the population, being able to prioritise the villages most at risk for public awareness campaigns in their organisation’s yearly operational plan
• Health policy-makers being able to incorporate into their multi-year strategic planning insights around how climate change, urbanisation and migration are predicted to affect the risk of snakebite envenoming and corresponding resource requirements in the region

Success indicators

**Impact indicator:** Reduced numbers of snakebites (due to better preventative activities), and reduced mortality and morbidity through snakebites (due to better resourcing/availability of anti-venom and other resources needed to treat snakebites).

**Mid-term outcome indicator:** Use of products and tools created during the challenge. Examples include informing healthcare decision-makers about antivenom requirements in a particular region or helping NGO leaders prioritise preventative activities for snakebite in Tamil Nadu.

Additionally to these success indicators, this data challenge will also play a key role in:

• **Increasing public trust:** as all insights and tools will be based on early engagement with relevant communities in the countries, and will be shared widely
• **Increasing capacity:** as training and mentorship will happen throughout the data challenge programme, making sure participants have the capacity to take part and build long term data science capacity that will benefit the country.
Potential time to impact

There is a potential for impact to be evident in a short period of time for this challenge topic, given the current lack of models and applications in this area in India, and the fact that snake bites occur suddenly and can lead to swift mortality.

Strategic alignment with Wellcome

Wellcome have made snakebite a priority area with the goal of helping transform the way in which snakebite treatments are researched and delivered. Wellcome has three main goals in this space: bringing the production of snakebite treatments into the 21st century; developing the next generation of treatments; and building and sustaining snakebite as a global health priority. Wellcome has a dedicated team committed to these ambitions and plan to achieve this by:

- modernising antivenom production by enhancing the science and applying technology to make products better, safer and cheaper
- promoting the development of new snakebite treatments and antivenoms by generating an evidence base for which treatments work and why
- generating evidence about snakebite's prevalence and economic cost to underpin the development of a sustainable market for treatments

By supporting the development of models and tools to improve understanding and response of snakebite that could scale across South Asia and to Sub-Saharan Africa, this challenge is directly aligned with Wellcome’s priorities.

Data landscape summary

There has been historic underreporting of snakebite cases and a paucity of data, but improvements in the quantity and quality of snakebite data make it possible to deliver meaningful insights for the healthcare system. In India, data exists from independent research, national surveys, snake catching programmes, social media and crowdsourced projects. These can provide information and insights on:

- snakebite incidence (including circumstances of the bite, e.g. type of snake, time of bite, characteristics of victim), including livestock incidence data
- seasonal and geographic variation in snakebite incidence
- snake populations (their species and preferred habitat, snake distribution - venomous and non-venomous)
- types of venom by species and location
- human population distribution
- sociodemographic and occupational profile of the population
- weather and climate patterns
- urbanisation and other manmade changes to the landscape
- healthcare facilities, antivenom storage and networks
- physical access to healthcare facilities, i.e. time needed to reach
- antivenom potency, quality and composition

Key to the challenge will be the patient incidence data from healthcare facilities and community survey data on snakebite instances. The accuracy and the consistency of the former varies
across and within states, but Tamil Nadu has been suggested as a state where the data infrastructure in place would allow for demonstrating what can be achieved through applying data science techniques to existing data.

There is also a research proposal currently at the peer review stage looking to do a systematic review on the burden of snakebite in India. The protocol outlines a systematic review and meta-analysis on the burden (incidence/prevalence, mortality, morbidity, health facility, and economic) of snakebite in India nationally and sub-nationally, and on the risk factors related to snakebite (bite, envenomation, death, adverse outcomes) in India. A challenge approach will provide an opportunity for sharing data across the multiple ongoing research projects. Connecting with researchers in adjacent fields to snakebite and herpetology, such as ecology and environmental science will improve the quality and range of potential outputs from the challenge. More information on key data, sources, and availability is in Data Inventory section.

Challenge structure

The challenge period will cover a period of 18 months, with a 9 month set up phase preceding the prize, and 9 month sustainability prize following the judging.

Seed funding would be needed in order to allow for people to form multi-disciplinary teams and take part in the competition. Details of this are available in the budget and in the descriptions below.

Training and mentoring would also be required, to align capacity and to ensure benefits would be lasting for the participants and the country as a whole. Specific training requirements will be established early in the challenge, depending on what participant needs are. Training might cover relevant areas from the ODI data skills framework, such as ethical use of data or machine learning approaches.

Remit and scope of the challenge

Key questions

- How can we help hospitals and other healthcare providers better plan and manage the resources needed to respond to snakebite instances (antivenom and other supplies and equipment (e.g. intensive care facilities), trained clinicians, etc.)?
- How can we help NGOs and other stakeholders better target snakebite prevention activities, e.g. community education interventions, based on short-term snakebite activity predictions?
- How can we help health policy-makers carry out better long-term planning for snakebite management requirements in the region based on long-term snakebite activity predictions?

Suitable proposals

Proposals should be submitted to make progress in one or more of the following areas:

---

63 https://theodi.org/article/open-data-skills-framework/
● Tools / applications that can inform public health officials in charge of antivenom distribution and coordination and the distribution of other resources needed to treat snakebite envenoming.
● Tools / applications that can be used by NGO leaders and others involved in snakebite prevention to improve prioritisation of community outreach programmes to educate the population on snakebite prevention and treatment
● Tools / applications that support health policy-maker resource planning around the effects of climate change, urbanisation and migration on snakebite envenoming

Any proposal submitted should consider the complexity of the snakebite process of a human, snake and environment interaction. The end goal is to inform healthcare planning, i.e. to help the optimisation and distribution of resources for prevention and treatment. This might include the optimal allocation of antivenom and intensive care units at the local level in Tamil Nadu.

Who can apply
We are looking for multidisciplinary teams who bring data science together with a range of alternative disciplines (such as herpetology, biology, zoology, ecology, environmental science, etc) in order to deliver innovative solutions to this complex issue.

The prize will be open to all international participants, but there will be significant engagement and recruiting with local networks of experts and all events will be run in the location specific to the prize.

The application process is open to those who have predefined ideas and projects, and those who may have relevant skills but no project proposal.

Participant motivations
The motivation for participants to enter will be a combination of:
● opportunities for networking / building relationships
  ○ access to key government and healthcare sector stakeholders
  ○ access to other experts within and across disciplines
  ○ access to commercial partners and clients
● personal and professional development
  ○ advancing the field of snakebite research
  ○ advancing technological development
  ○ advancing individual capabilities
● personal and professional rewards
  ○ financial reward of winning the prize
  ○ reputational reward of winning the prize

The motivation for academic participants tends to be more driven by access to key government and healthcare stakeholders, improvement in and contributions to their field of study, and reputational awards.

The motivation for commercial participants tends to be more driven by access to new markets and clients, positive press around service provision and financial rewards.
The motivation for governmental stakeholders will be to outsource the development of innovations that can help them deliver more efficient services, and contribute towards SDG goals.
Challenge stages, funding & support

The estimated time allocations for the snakebite data challenge would include:

- **Up to 9 months** for a set up phase
- **Up to 18 months** for the challenge to run, from recruitment of participants through to the awarding of the challenge prizes
- **Up to an additional 9 months** for a sustainability period

For a detailed view of estimated timelines, please refer to the snakebite data challenge delivery plan in Appendix 10.

1. **Set up phase**
   
   The organisation that is selected to run this challenge programme would likely need six to nine months in order to set up the challenge programme. This might include, but is not limited to, the following activities:
   
   1. Convening end-users and stakeholders across the communities in question, local, state and national government and private sector in order to verify their needs with regard to snakebite and confirm their engagement with the prize.
   2. Establish agreements with local delivery partners, and establish the detailed delivery plan.
   3. Establish agreements with data providers.
   4. All data will need to be vetted and cleaned thoroughly prior to launch in order to verify quality and limitations. This will take dedicated resources and be the most time consuming part. The aim is to ensure that participants can use the data immediately at the point they enter the Model Development Phase.

2. **Participant recruitment/ Expression of interest**

   A period of up to two months will be required for recruitment of participants. This would run concurrently with the set up phase. A website will need to be established with all the required information on topic and processes. This will then be publicised through some advertising and media channels. In this case, as it is likely that the bulk of the participants will be recruited through established networks in the snakebite focus area, participants will be invited to complete an expression of interest (EOI).

   EOI can be open to those who have predefined ideas and projects, and those who may have relevant skills but no project proposal. Participants will be able to
communicate with each other during this period, and will be encouraged to make connections and discuss ideas to help with team formation.

3. **Launch lab**  
The challenge prize will begin with a ‘lab’ or multi-day event (possibly virtual given the Covid-19 crisis). All those who have completed an EOI will be invited, along with beneficiaries, data providers and delivery partners. Participants will have their expenses paid for travel and accommodation.

The first part of this event will focus on explaining why this challenge topic is so important, telling the stories of why and how this problem needs to be addressed, and the processes involved in participating in the challenge prize. This can be recorded to be sent to people who cannot attend. Press will be invited to cover the ‘launch’ and meet some of the participants and key stakeholders.

Following this introductory session, participants will have the opportunity to share their project ideas, recruit new team members, and engage with data providers and stakeholders.

The aim is to support the development of multidisciplinary teams, bring them into contact with the data providers and beneficiaries, and kick start the development of their proposals and ideas through two days of workshops and activities. There will be a clear process for how to team up and formally enter as a team.

4. **Team applications**  
The teams will be given 4-8 weeks following the launch lab to refine their proposal ideas, confirm their teams and submit their area of interest/ initial proposal in a formal team application. This will then be assessed based on criteria including:
   - Multidisciplinary team with relevant expertise  
   - Area of interest in line with challenge scope  
   - Scale of ambition

During this period, participants will have access to a challenge team application template to complete, have access to supporting information on the topic, and will be able to troubleshoot with the organisation running the challenge for process queries.

Teams may submit their proposal at any time during this period and begin the next phase, though there will not be access to funding before 4 weeks. The application window will remain open until 8 weeks after the launch lab.

5. **Model development phase**  
This phase will be between 6-9 months and focus on training and the development of data insights. The teams will be encouraged to focus on engaging with the available data and modeling it to develop insights. During this time they will receive some seed funding (£25k for up to 10 teams), data access support to help them identify the specific data sets they need for their project and any governance requirements, and a rolling series of seminars on relevant machine learning techniques, public health and other relevant topics.
In this phase we would expect the members of the teams with data science, public health and environmental science to be leading the work.

Through the development of models and insights, the teams will be asked to put together plans for developing applications and tools drawing on the data science methods and insights they have developed in the model development phase. As stated in the challenge scope, these tools and applications will need to deliver specific benefits such as, but not limited to:

a. Informing public health officials managing antivenom distribution and coordination and the distribution of other resources necessary to treat snakebites.
b. Improving prioritisation of community outreach programmes for education on snakebite prevention and treatment
c. Supporting health policy-maker resource planning regarding climate change, urbanisation and migration on snakebite envenoming

The model development phase will culminate in a demo day in which all of the teams will demo the models and insights they have developed using the data, and their pitches for tool development. These will be judged by a panel of key stakeholders and experts in the field against a set publicly accessible scoring criteria. For example, applications might be assessed on some of the following areas:

- Understanding of the problem
- The proposed approach
- Team resources
- Proposed outcomes
- Potential impacts
- Project management documentation
- Risk assessment
- Cost plan
- Ethical assessment

The best teams will be selected to go through to the development stage to build their proposed tools/applications.

6. Tool development phase
The tool development phase will also run for between 6-9 months. The projects that have progressed will receive further seed funding (£100k for up to 5 teams) in order to develop the tool or application outlined in the model development phase.

In this phase we would expect the teams to have a greater focus on digital development, working with and drawing on the expertise of the scientific experts and stakeholders to inform their work. The challenge teams will get a chance to implement their full or partial solutions in an environment with real stakeholders. This will allow for feedback, adjustments and further development before the final judging. These teams will receive specific training and mentorship suitable to their specific topic to enable their ideas to develop and flourish. This will include support to plan implementation with stakeholders, ethical assessments and business model development.
7. **Judging**

The development phases will culminate in a public demo day attended by as many media representatives and senior stakeholders attending as possible. Each of the teams present their work. The event will have a strong focus on storytelling - what the nature and scale of the risks around urban public health are, and how we have used data science to address them.

A set of judges who are known experts in the field, such as Romulus Whitaker, François Chappuis and David Williams, will review the submissions against publicly available criteria. Criteria are likely to include:

- **Innovation**
  - Ethical data management
  - Equitable solution
  - Utilises multiple data sources

- **Impact**
  - Scale of potential impact on snakebite related health outcomes in the region
  - Clear plans on how the solution could improve trust in the relevant health ecosystem

- **Capacity**
  - Clear input from multiple disciplines

- **Beneficiaries**
  - Clear link to government/other appropriate end user
  - Scalable solution which has the potential to be used in other regions

- **Sustainability**
  - A clear, time bound implementation plan for the solutions to be adopted by end users
  - Whether the solution have the potential to continue beyond the prize

The most successful submission will be awarded a prize of £200k. Three runner up teams shall receive £100k. This money will be explicitly for the implementation and scaling of the tool they have developed. The rationale for this model is that it allows multiple tools and outputs to go on to be scaled and supported, delivering sustainable impact in the field. This has been benchmarked against other health technology innovation prizes⁶⁴.

8. **Sustainability / Uptake period**

It is vital that there is a sustainability period following the award of the prize in order to support the winning project and the other projects to be adopted by the key stakeholders and used to address the key issues addressed regarding mental health. The winning teams can use the prize money to implement and scale up the innovations they have developed.

In addition, the partners involved in the delivery of the prize will need to remain available for the teams over a period of 9 month in order to:

---

⁶⁴ Appendix 9
• Assist the future development of winning solutions by making introductions to would–be users, investors, commissioners and others.
• Establish a schedule for checking progress with winners and interesting other participants.
• Seek feedback from competitors, judges and partners and evaluate the effectiveness of the programme.
• Return to the prize winners a year afterwards to develop a case study of where they have got to in their development and the impact they have created.

9. Evaluation
At the end of the judging and sustainability phase there shall be an independent evaluation made of the prize as a whole drawing on reporting material and interviews with participants and stakeholders. The interviews will also be used to develop case studies of each output and prize in order to share learnings and successes for others interested in the field of health data innovation and remote sensing for public health.

Challenge prize data ecosystem
An ecosystem map can be helpful in understanding how data flows between different stakeholders involved in snakebite management and how value is created in the ecosystem by allowing different stakeholders to make better decisions.

To produce the below diagram, we used the ODI’s data ecosystem mapping tool. It is a visual illustration of how data is being accessed, used and shared.

Snakebite Challenge Data Ecosystem Map. A high res version of the map is available in Appendix 12 and an interactive version is available here.
This data ecosystem map displays: the potential data flows to challenge participants addressing snakebite; the potential value flows to the healthcare and research communities from the challenge participants; and the data and value flows underpinning antivenom distribution.

The challenge would aim to establish interdisciplinary partnerships, including for data provision, challenge co-design, advisory, etc.

**Risks**

The key risks related to the snakebite data challenge prize are primarily concerned with engagement and implementation with the healthcare sector, and the capability of challenge participants. If the healthcare system is unwilling to grant access to data, or does not implement useful tools that are created then the level of snakebite mortality and morbidity will not improve. Conversely, if capable multidisciplinary teams are unable to form or create meaningful tools, the desired improvements will not be attained.

The challenges teams, alongside their partners and the Wellcome trust may need to work together to overcome the potential unwillingness from the healthcare system to engage with the team and/or implement the tools. This could come in the form of providing helpful demonstrations and/or testimonies of similar tools in different sectors or regions, providing financial or technical assistance, or tying engagement and/or implementation to other commitments. The team can also build on the already established relationships of partners, such as the Madras Crocodile Bank Trust, who work closely with the Department of Health and can help engage with key stakeholders in the healthcare sector.

A full risk register is available in Appendix 3.

**Capacity and stakeholder landscape**

**Data science skills**

India has a vibrant data science community. According to the Analytics Report 2018 by Analytics Peepal, India is “expected to be in the top 3 big data analytics markets by 2020”. By then, the industry is “expected to almost double, with 24% being attributed to Big Data and 12% to advance analytics, predictive modeling and data science.” As per a study run by AIM & Praxis Business School in 2019, the analytics, data science and big data industry in India is expected to double by 2025. Both sources place Bengaluru and Delhi NCR as the two biggest contributors to the market size in analytics, together capturing just over half the market. Mumbai, Pune, Hyderabad and Chennai follow these two cities in the list of analytics hubs according to the AIM & Praxis Business School study.

---

65 Analytics Report 2018, Analytics Peepal
66 Analytics & Data Science Industry In India: Study 2019 By AIM & Praxis Business School
Besides multiple leading universities and colleges offering data science programmes, research institutes such as the Wadhwani Institute for Artificial Intelligence that focuses on developing AI solutions for social good offer opportunities for collaboration.

Healthcare service providers

In 2019, Tamil Nadu ranked second, following the state of Kerala, in the National Institute for Transforming India’s (Niti Aayog) health index report, and the state capital is often referred to as India’s health capital. Major healthcare providers in the state include The Tamil Nadu Dr. M.G.R. Medical University (TNMGRMU), Rajiv Gandhi Government General Hospital, Tamil Nadu Government Multi Super Specialty Hospital, Stanley Medical College (SMC) and Christian Medical College & Hospital, Vellore (CMC,Vellore).

One of the potential key beneficiaries of the challenge is the Tamil Nadu Medical Services Corporation (TNMSC). TNMSC is the sole procurement organisation for medical equipment and drugs in the state of Tamil Nadu, therefore it is the single largest buyer of antivenoms in the state, and the single largest direct supplier of antivenoms to healthcare providers. Antivenoms are supplied to healthcare providers based on their estimated requirements for a multi month period.

Researchers, third sector and industry

Organisations such as the Madras Crocodile Bank Trust and Centre for Herpetology based in Tamil Nadu, the George Institute for Global Health with offices in New Delhi and Hyderabad, the Snakebite Healing and Education Society and the Indian Institute of Science are actively involved in snakebite mitigation efforts and are interested in better using data to reduce the snakebite burden in India. They could likely get involved or know people that could get involved as applicants to the challenge. These actors are also well-connected to the healthcare system at both state and national levels.

The DBT/Wellcome Trust India Alliance (India Alliance) is a public charity that funds research in health and biomedical sciences in India, itself funded by the Department of Biotechnology, Government of India and the Wellcome Trust. India Alliance has a history of running health related challenges and grants in India, and given the relationship with Wellcome may be a useful organisation to consult in the set-up of the challenge both to seek input for the design of the challenge as well as to help promote the challenge prize among their networks.

72 India Alliance, “Welcome”, https://www.indiaalliance.org/
This is the list of key organisations to be considered and consulted with in delivering this challenge:

| Research / Academia / Activism                                                                 | • Liverpool School of Tropical Medicine (LSTM)  
|                                                                                               | • University of Geneva / Geneva University Hospitals (UNIGE/HUGs)  
|                                                                                               | • Madras Crocodile Bank Trust (MCBT)  
|                                                                                               | • University of Copenhagen (UCPH)  
|                                                                                               | • Snakebite Healing and Education Society (SHE Society)  
|                                                                                               | • The George Institute for Global Health  
|                                                                                               | • Imperial College London  
|                                                                                               | • University of Kelaniya, Sri Lanka  
| Technology Institutes                                                                        | • Indian Institute of Science (IISc)  
|                                                                                               | • Vellore Institute of Technology (VIT) and VIT Chennai  
| Tamil Nadu Healthcare                                                                        | • The Tamil Nadu Dr. M.G.R. Medical University (TNMGRMU)  
|                                                                                               | • Rajiv Gandhi Government General Hospital  
|                                                                                               | • Madras Medical College  
|                                                                                               | • Tamil Nadu Government Multi Super Speciality Hospital  
|                                                                                               | • Government Kilpauk Medical College (GKMC)  
|                                                                                               | • Government Royapettah Hospital  
|                                                                                               | • Stanley Medical College (SMC)  
|                                                                                               | • Christian Medical College & Hospital, Vellore (CMC,Vellore)  
| Co-Funders                                                                                   | • Wellcome Snakebite Priority Area  
|                                                                                               | • Bill & Melinda Gates Foundation (BMGF)  
|                                                                                               | • Tata Trust  
|                                                                                               | • Ford Foundation  
|                                                                                               | • Rockefeller Foundation  
|                                                                                               | • The DBT/Wellcome Trust India Alliance (India Alliance)  
| International bodies                                                                        | • World Health Organization (WHO)  
|                                                                                               | • United Nations (via SDGs)  
| Telecoms organisations                                                                       | • International Telecommunication Union (ITU), Geneva  
|                                                                                               | • GSMA, London  
|                                                                                               | • Indian telecoms providers (Jio, Vodafone, Airtel, BSNL)  
| Government and Regulators                                                                    | • Indian Council of Medical Research / Indian Medical Council  
|                                                                                               | • National Electronic Health Authority |
An overview of the different stakeholders involved in snakebite management is available in the below stakeholder map developed by a design and innovation agency based in India.

**Snakebite Stakeholder Map, Quicksand**

![Stakeholders in the Snakebite Ecosystem](image)


**Wider Challenge Landscape**

**WHO Snakebite Roadmap**

The aim of the WHO Snakebite strategy is to halve the numbers of deaths and cases of disability due to snakebite envenomation by 2030 through a programme that targets affected communities and their health systems, and by ensuring access to safe, effective treatment.

---

through increased cooperation, collaboration and partnership at all levels. The cost of implementing the four strategic objectives between 2019 and 2030 will be spread over three phases:

- Pilot Phase (2019-2020) 10–12 countries US$ 8.96 million
- Scale-up Phase (2021-2024) +35–40 countries US$ 45.44 million
- Full Roll-out (2025-2030) All affected countries US$ 82.36 million

The programme has four strategic aims:

1. Empower and engage communities - Prevent snakebite envenoming and increase use of treatment through education, training and facilitation.
2. Ensure safe, effective treatment - Build a stable, sustainable market for safe, effective antivenoms at reasonable cost and assured access to treatment.
3. Strengthen health systems - The principles of the WHO Health Systems Framework should be used to integrate more effective prevention, treatment and management of snakebite envenoming into national health systems, national health plans and policy frameworks.
4. Increase partnerships, coordination and resources - Strong collaboration will be required for this comprehensive plan of action, and advocacy will be necessary to build a global coalition to drive change, generate investment, implement projects and accelerate research into new therapies, diagnostics and medical interventions.

**Indian Snakebite Initiative: Big4 Mapping**

Big4 Mapping by the Indian Snakebite Initiative is a collaborative, national level citizen science project with the aim of mapping snake distribution and snakebite incidence of the four medically important snakes of India (common krait, russell’s viper, indian saw-scaled viper, indian cobra) which cause more than 95% of snakebite deaths. The mapping - the first ever extensive mapping of the medically important snakes across the country- is done by more than 800 volunteers using the Big-4 mobile app across India. The project started in 2017 May under the guidance of the Madras Crocodile Bank Trust.

Researchers, funded in part by Wellcome, from the University of Geneva have begun using the data generated from this project and others to develop artificial intelligence / machine learning (AL / ML) solutions to identify snakes, and therefore prescribe snake antivenom, faster.²⁵

**National initiatives in other countries**

Costa Rica has taken a similar approach to map the national burden of snakebite that is proposed in this report. Costa Rica has been able to manage snakebite locally quite well due to more accessible healthcare and antivenom services, fewer poisonous snakes, relatively small area and low prevalence of traditional healing methods for snakebite. They have used geographical information system (GIS)-powered solutions increasingly as a tool for improved decision making regarding snakebite management. In 2013 a research team published a study to provide information to assist decision-making concerning which primary health care facilities

---


in Costa Rica should have antivenom, by understanding which serve a population with a high risk of snakebites and long transport times to hospitals or clinics where antivenom is available.76

India’s neighbouring states of Nepal and Sri Lanka face similar issues to India but on a smaller geographic and demographic scale. In Nepal, nationwide community surveys have been used to understand the true burden of snakebite and understand its geographic distribution. Survey data showed a difference to official numbers by an order of magnitude.77 The Nepalese healthcare system is now responding to this new information by changing antivenom distribution from primarily hospital-centric to mobile deployment. In Sri Lanka, a similar geospatial mapping project to Costa Rica was delivered which examined the risk of snakebites and evenoming on a national scale. Their modelling also included land use and climatic zones as variables. Another project in Sri Lanka, not yet published, combined land use change and climate change modelling with snakebite and is expected to lead to a stronger analysis and higher resolution. In addition, an agent-based modelling approach is being applied, using farmer activity as a way to predict snakebite patterns going forward.78 Other countries, such as Iran are also using seasonal time series methods to better predict future cases of snakebites and help public health authorities design necessary measures. 79

Legal, policy and regulatory environment

This is a summary of the ‘Health and agriculture overview’ report commissioned to the Aapti Institute in India, the full report can be found in Appendix 16.

While there are several proposed protections, laws and regulations, India currently does not have a clear data protection framework and challenge teams working with health data will likely face a complex legal landscape with overlapping laws and regulations. Privacy is a fundamental right in India, and the most extensive piece of legislation protecting this right is the Personal Data Protection Bill, 2019 (PDP 2019), which has yet to pass Indian parliament. The COVID-19 pandemic is currently delaying passage of this critical bill, however, legal, policy and compliance experts in the country are very much keeping abreast of developments. Data challenge participants should be able to find the right expertise to ensure they are complying with PDP 2019 and the data regulation landscape.

Other key policies regarding digital healthcare and data access include the National Health Policy, 2017, the Digital Information Security in Healthcare Act (DiSHA), and the National Data Sharing and Accessibility Policy (NBSDAP). Ideally with the passage of PDP 2019 a mandated Data Protection Authority (DPA) will be able to provide a harmonised view of the regulatory landscape to individuals and organisations seeking clarity on compliance.

78 ibid
Data sharing is clearly a priority for the government. The development of the Open Government Data (OGD) platform which contains 221,300 data sets in 730 catalogues across central and state government activities shows the commitment of the government and is key data infrastructure for any health data challenge. This health data challenge represents an opportunity to demonstrate the opportunity from greater secure data sharing.

However, given the ever changing policy landscape, ongoing professional legal support will be required to ensure challenge participants are able to access important and potentially sensitive and personal health data, as well as use it to create tools for the challenge. As of May 2020, it is uncertain when PDP 2019 will be passed or in exactly what form, but it may happen just before or even during the challenge. For this reason there needs to be a budget for consistent legal support for participants throughout the challenge.

In addition to the changing legislative context, there is at present low capacity within the health sector and lower levels of government around data management and the uptake of data innovations like the current challenge prize. Funding a sustainability phase will support uptake and implementation of the tools designed in the challenge and may present opportunities for further funding within the Data for Science and Health programme by building scalable models across the country, or models that could be adapted for use in other areas of public health.

Data Inventory

The below table is a high level analysis of relevant data sources and their availability to potential challenge participants. The government of India has made a significant amount of relevant data sources available through its open data portal. Other datasets are held by national and international research institutes with a range of accessibility, from not being shared beyond the institution itself, to openly published. A key challenge will be access to healthcare data from healthcare providers, as well as the quality and scope of that data.

<table>
<thead>
<tr>
<th>Data type</th>
<th>Organisation/ Source</th>
<th>Availability</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agricultural land</td>
<td>PREP Data / Bhuvan, Indian Space Research Organisation (ISRO), National Remote Sensing Centre (NRSC)</td>
<td>Internal access</td>
<td>Land Use / Land Cover (LULC) for India on 1:250,000 (30 m) scale, including built-up, agricultural, and natural areas.</td>
</tr>
<tr>
<td>Ambulance</td>
<td>Open Government Data Platform India</td>
<td>Open</td>
<td>State/UT-wise Ambulances Operational under National Health Mission (NHM) during 2018</td>
</tr>
<tr>
<td>Antivenom</td>
<td>characteristics</td>
<td>Antivenom producers</td>
<td>Public</td>
</tr>
<tr>
<td>Antivenom stocks</td>
<td>Healthcare providers</td>
<td>Missing /</td>
<td>Stock of antivenom supplies held by clinics and other healthcare providers</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Privately held</td>
<td></td>
</tr>
<tr>
<td>Climate Change</td>
<td>IRI/LDEO Climate Data Library</td>
<td>Public</td>
<td>Climate trends showing how wet and dry (or hot and cold) periods have varied over the past century, including Year-to-year and decade-long shifts and century-long trends</td>
</tr>
<tr>
<td>Category</td>
<td>Source</td>
<td>Accessibility</td>
<td>Notes</td>
</tr>
<tr>
<td>--------------------------------</td>
<td>------------------------------------------------------------------------</td>
<td>---------------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Construction</td>
<td>Open Government Data Platform India</td>
<td>Open</td>
<td>Note: there are various construction datasets published on the India Open Government Data Portal that may be useful to the challenge team.</td>
</tr>
<tr>
<td>Ecology</td>
<td>Imperial College London</td>
<td>Accessible to researchers</td>
<td>Data on snake ecology collected by Imperial College London.</td>
</tr>
<tr>
<td>Elevation</td>
<td>DIVA-GIS</td>
<td>Public</td>
<td>Free geographic (GIS) data, including elevation, for any country in the world.</td>
</tr>
<tr>
<td>Forestation</td>
<td>Open Government Data Platform India</td>
<td>Open</td>
<td>Data refers to the Forest Cover of India, all lands more than one hectare in area, with a tree canopy density of more than 10 percent irrespective of ownership and legal status.</td>
</tr>
<tr>
<td>Healthcare facility</td>
<td>OpenStreetMap</td>
<td>Open</td>
<td>Geospatial data about healthcare facilities across the state.</td>
</tr>
<tr>
<td>Healthcare facility</td>
<td>Facebook Disease Prevention Maps</td>
<td>Accessible for nonprofit</td>
<td>Facebook Disease Prevention Maps are designed to help public health organizations close gaps in understanding where people live, how people are moving, and the state of their cellular connectivity, in order to improve the effectiveness of health campaigns and epidemic response.</td>
</tr>
<tr>
<td>Healthcare facility</td>
<td>Healthsites.io</td>
<td>Public</td>
<td>Geospatial data about healthcare facilities across the state.</td>
</tr>
<tr>
<td>Healthcare Infrastructure in Rural Areas</td>
<td>Open Government Data Platform India</td>
<td>Open</td>
<td>State-wise data about number of administrative units and rural health infrastructure (DHs, CHCs, PHCs and Sub-Centres).</td>
</tr>
<tr>
<td>Herpetology</td>
<td>The Reptile Database</td>
<td>Public</td>
<td>Public database on reptile species.</td>
</tr>
<tr>
<td>Herpetology</td>
<td>PLOS neglected tropical diseases</td>
<td>Open</td>
<td>Geographical distribution of MIVS citizen-generated and scientist-generated observation data.</td>
</tr>
<tr>
<td>Herpetology</td>
<td>University of Geneva - Open Archive</td>
<td>Open</td>
<td>Research data on snakebite identification.</td>
</tr>
<tr>
<td>Herpetology</td>
<td>University of Geneva - Open Archive</td>
<td>Open</td>
<td>Research data on snakebite identification.</td>
</tr>
<tr>
<td>Household by building material</td>
<td>Directorate of Census Operations Tamil Nadu / Census India</td>
<td>Open</td>
<td>District breakdowns of number of households by material used for roof, walls, floors, etc.</td>
</tr>
<tr>
<td>Household by cooking fuel</td>
<td>Directorate of Census Operations Tamil Nadu / Census India</td>
<td>Open</td>
<td>District breakdowns of number of households by cooking fuel.</td>
</tr>
<tr>
<td>Household by latrine type</td>
<td>Directorate of Census Operations Tamil Nadu / Census India</td>
<td>Open</td>
<td>District breakdowns of number of households by latrine type.</td>
</tr>
<tr>
<td>Migration</td>
<td>Directorate of Census</td>
<td>Open</td>
<td>Census of India 2001 data highlights.</td>
</tr>
<tr>
<td>Data Type</td>
<td>Source</td>
<td>Privacy Level</td>
<td>Description</td>
</tr>
<tr>
<td>---------------------------------</td>
<td>------------------------------------------------------------------------</td>
<td>---------------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Mobile location data</td>
<td>Operations Tamil Nadu / Census India</td>
<td>Private</td>
<td>Location-based data derived from mobile device usage</td>
</tr>
<tr>
<td>Mobile location data</td>
<td>Telecoms providers (Vodafone India, Airtel, BSNL, Jio)</td>
<td>Private</td>
<td>Location-based data derived from mobile device usage</td>
</tr>
<tr>
<td>Natural disasters</td>
<td>Our World in Data</td>
<td>Open</td>
<td>Temporal data of natural disasters over time</td>
</tr>
<tr>
<td>Number of Public Healthcare</td>
<td>Open Government Data Platform India</td>
<td>Open</td>
<td>State/UT-wise details of the Number of Public Healthcare Centers with Trained Providers</td>
</tr>
<tr>
<td>Centers with Trained Providers</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other demographic data</td>
<td>Directorate of Census Operations Tamil Nadu / Census India</td>
<td>Open</td>
<td>Note: there are additional demographic datasets published by Census India that may be useful to the challenge team</td>
</tr>
<tr>
<td>Other environment data</td>
<td>Open Government Data Platform India</td>
<td>Open</td>
<td>Note: there are additional environment datasets published on the India Open Government Data Portal that may be useful to the challenge team</td>
</tr>
<tr>
<td>Other healthcare data</td>
<td>Open Government Data Platform India</td>
<td>Open</td>
<td>Note: there are additional healthcare datasets published on the India Open Government Data Portal that may be useful to the challenge team</td>
</tr>
<tr>
<td>Population</td>
<td>Directorate of Census Operations Tamil Nadu / Census India</td>
<td>Open</td>
<td>Results from the 2001 India census, including population and population density of Tamil Nadu</td>
</tr>
<tr>
<td>Population density</td>
<td>Directorate of Census Operations Tamil Nadu</td>
<td>Open</td>
<td>Results from the 2001 India census, including population and population density of Tamil Nadu</td>
</tr>
<tr>
<td>Railroads</td>
<td>DIVA-GIS</td>
<td>Public</td>
<td>Free geographic (GIS) data, including railroads, for any country in the world</td>
</tr>
<tr>
<td>Rainfall</td>
<td>Department of Climate and Rainfall, Government of Tamil Nadu: Hydromet division of the Indian Meteorological Department</td>
<td>Open</td>
<td>Data about Annual Rainfall in Tamil Nadu Subdivision during 1901 to 2017</td>
</tr>
<tr>
<td>Recently rezoned land</td>
<td>TBD</td>
<td>Missing</td>
<td>Data about changes in land use, specifically from natural to agricultural land</td>
</tr>
<tr>
<td>Roads</td>
<td>DIVA-GIS</td>
<td>Public</td>
<td>Free geographic (GIS) data, including roads, for any country in the world</td>
</tr>
<tr>
<td>Category</td>
<td>Source/Description</td>
<td>Accessibility</td>
<td>Notes</td>
</tr>
<tr>
<td>----------------------------------</td>
<td>----------------------------------------------------------</td>
<td>---------------</td>
<td>----------------------------------------------------------------------</td>
</tr>
<tr>
<td>Snake observation</td>
<td><strong>Snakebite initiative - Big 4 Mapping</strong></td>
<td>Private</td>
<td>Citizen-generated geospatial data about observed instances of snakes</td>
</tr>
<tr>
<td>Snake observation</td>
<td><strong>Snake Identification Service Sri Lanka</strong></td>
<td>Private</td>
<td>Citizen-generated geospatial data about observed instances of snakes</td>
</tr>
<tr>
<td>Snake observation</td>
<td><strong>Citizen Science snake ID data set</strong></td>
<td>Private</td>
<td>Citizen-generated geospatial data about observed instances of snakes</td>
</tr>
<tr>
<td>Snakebite compensation claims</td>
<td>Local authorities</td>
<td>Missing</td>
<td>Snakebite is a compensable event in India, local authorities may have data on snakebite that healthcare providers do not</td>
</tr>
<tr>
<td>Snakebite incidence</td>
<td><strong>Madras Crocodile Bank Trust</strong></td>
<td>Private</td>
<td>Data compiled from healthcare institutions in Tamil Nadu on snakebite incidence</td>
</tr>
<tr>
<td>Snakebite incidence</td>
<td><strong>Department of Health and Family Welfare of State of Tamil Nadu</strong></td>
<td>Private</td>
<td>Data across various healthcare institutions in Tamil Nadu on snakebite incidence</td>
</tr>
<tr>
<td>Snakebite incidence</td>
<td><strong>PLOS Neglected Tropical Diseases</strong></td>
<td>Open</td>
<td>Data from a nationally representative mortality survey on snakebite in India from 2011</td>
</tr>
<tr>
<td>Snakebite incidence</td>
<td><strong>PLOS Neglected Tropical Diseases</strong></td>
<td>Open</td>
<td>Data from a household survey (28,494 people) of snakebite incidence combined with a more detailed survey of victims in order to understand the health and socio-economic effects of the bite, the treatments obtained and their views about future improvements.</td>
</tr>
<tr>
<td>Snakebite incidence</td>
<td><strong>Indian Council of Medical Research</strong></td>
<td>Private</td>
<td>Data across various healthcare institutions across India on snakebite incidence</td>
</tr>
<tr>
<td>Snakebite, envenoming, mortality</td>
<td><strong>Global snakebite registry</strong></td>
<td>Private</td>
<td>(under development) Data portal by snakebite researchers combining international sources</td>
</tr>
<tr>
<td>Snakebite incident reports</td>
<td>Police stations</td>
<td>Missing</td>
<td>Snakebite is not assumed to be an accident in India, local police therefore may have data on snakebite that healthcare providers do not</td>
</tr>
<tr>
<td>Specific humidity</td>
<td><strong>IRI/LDEO Climate Data Library</strong></td>
<td>Public</td>
<td>Data on daily intrinsic pressure level for humidity over time</td>
</tr>
<tr>
<td>Sunlight hours</td>
<td><strong>IRI/LDEO Climate Data Library</strong></td>
<td>Public</td>
<td>Data on hours of sunlight per region over a solar year</td>
</tr>
<tr>
<td>Temperature</td>
<td><strong>IRI/LDEO Climate Data Library</strong></td>
<td>Public</td>
<td>Data on daily temperature over time</td>
</tr>
<tr>
<td>Urbanisation</td>
<td><strong>Census India</strong></td>
<td>Missing</td>
<td>Urbanisation trends by district for Tamil Nadu from 1991 to 2011</td>
</tr>
<tr>
<td>Vegetation</td>
<td><strong>Open Government Data Platform India</strong></td>
<td>Open</td>
<td>State/UT-wise mangrove cover assessment</td>
</tr>
<tr>
<td>Weather / Climate</td>
<td><strong>IRI/LDEO Climate Data Library</strong></td>
<td>Public</td>
<td>Online data repository and analysis tool that allows a user to view, analyze, and</td>
</tr>
<tr>
<td>Worker</td>
<td>Directorate of Census Operations Tamil Nadu / Census India</td>
<td>Open</td>
<td>download hundreds of terabytes of climate-related data through a standard web browse</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Data on employment breakdown in Tamil Nadu, including by sex, rural-urban, and other demographic factors</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Drug resistance infections (DRI) health data challenge: Detailed Scoping

**Covid-19**

Due to the wave of Covid-19 infections, in some countries considered below, we could anticipate that the consumption of antibiotics would go up significantly (people trying to treat initial symptoms without understanding that antibiotics are ineffective against a viral infection). Even though Covid-19 is a viral illness not affected by antibiotics, early data from hospitals shows that high proportions of patients are being treated with antibiotics to cure or protect against secondary infections during respiratory illnesses or hospitalization\textsuperscript{80}. There could also be issues around stocks, as some countries might hold on to their current stocks. In this context, the focus for a data challenge on drug resistant infections would need to be revisited.

The landscape and priorities could also potentially change due to a wave of locusts in East Africa, which presents "an extremely alarming and unprecedented threat" to food security and livelihoods, according to the United Nations\textsuperscript{81}.

---

**Summary recommendation**

**Key challenge question**

How can we use data science to optimise the use of antibiotics in order to minimise the local emergence and spread of antimicrobial resistance in Uganda, Malawi and Kenya?

**Why?**

Globally, antimicrobial resistance threatens the effective prevention and treatment of an increasing range of infections caused by bacteria, parasites, viruses and fungi. This has far reaching consequences. The UN estimates that more than 200,000 newborn children die each year from infections that do not respond to available antibiotics\textsuperscript{82}. More than 490,000 people developed multi-drug resistant tuberculosis globally in 2016, and drug resistance is complicating the fight against HIV and malaria\textsuperscript{83}. Each year, there are up to 700,000 deaths

---


\textsuperscript{82} Birth in a time of antibiotic-resistant bacteria (August 2016), WHO, [https://www.who.int/mediacentre/commentaries/antibiotic-resistant-bacteria/en/](https://www.who.int/mediacentre/commentaries/antibiotic-resistant-bacteria/en/)

\textsuperscript{83} Antimicrobial resistance (February 2016), WHO, [https://www.who.int/news-room/fact-sheets/detail/antimicrobial-resistance](https://www.who.int/news-room/fact-sheets/detail/antimicrobial-resistance)
globally due to drug resistance in illnesses such as bacterial infections, malaria, HIV/AIDS and tuberculosis\textsuperscript{64}.

Using antibiotics that do not correspond to your condition – or a wrong combination of them, for the wrong period of time – builds resistance. Therefore, understanding how antibiotics are being prescribed and how patients are accessing them is key to understanding and stopping the spread of resistance.

In the UK, access to antibiotics is regulated, people need to visit a doctor and get a prescription before accessing antibiotics in a pharmacy. However, that is not the case everywhere. In some low and middle income countries, many antibiotics are available without prescription. They can be bought over the counter, in unlicensed drug stores and in open vans in markets, for example. People buy the antibiotics they can afford, take a lot of different ones at the same time, and not for the right amount of time. This builds a lot of resistance and is almost invisible as there is so little data on the prescription and consumption of antibiotics through these informal dispensaries. These issues have been observed in Uganda, Malawi and Kenya.

What?

**How can we use data to optimise the use of antibiotics in order to minimise the local emergence and spread of antimicrobial resistance in Uganda, Malawi and Kenya?**

The objective of this challenge would be to run a data science challenge prize with a package of seed funding, training and sustainability support amounting to over £1.25million. This prize will bring together data collected through recent and upcoming initiatives (detailed in the annex on the wider challenge landscape), in order to improve where and how people are getting access to antibiotics and how they are using them:

- data on people’s behaviours related to antibiotic acquisition and consumption (availability, cost, use, prescribing and dispensing practices)
- epidemiological data (type of infection, treatment and outcome) and antibiotic susceptibility data

This objective would be in line with the national AMR plans detailed below as it would optimise the use of antimicrobial medicines in human health.

This challenge would make a difference through a new combination of data (behavioural and epidemiological data), non existing in current AMR initiatives, which would lead to new insights and drive better decision making. This could for instance take the shape of new community pharmacy maps and new analyses of antibiotics use, which would help prioritise interventions.

Currently, some key sources of data are either nonexistent (citizen generated and behavioural data to be collected via new initiatives detailed in the annex on the wider challenge landscape) or do not provide good quality data (a new initiative is aiming to collect better quality epidemiological and susceptibility data). This represents a key risk for this challenge as it relies on data currently or in the process of being collected and made accessible. This risk cannot be mitigated at this stage and will need to be closely monitored.

Where?

In the proposal below you will find specific information related to Uganda, Kenya & Malawi. Ghana has also been suggested, however it would need further investigation.

These countries have been chosen due to the fact that there are some existing relationships and some work being done on data collection through Wellcome’s DRI team and Wellcome’s public engagement team:

- The DRI team is working towards strengthening the DRI surveillance system in countries such as Kenya, Uganda, Ghana and Malawi, so they can collect comprehensive surveillance data. This data will provide an estimation of the burden of resistant infections and strains of resistant bacteria at the population level The first datasets ill be available in 2021.
- The public engagement team is working on community led initiatives in Malawi as well as on a potential citizens generated data initiative related to DRI in Kenya. This citizens generated data would provide key information on people’s behaviours towards antibiotics (how they access and consume them).

Many countries in sub-Saharan Africa suffer from a high burden of infectious diseases and challenges around access to medicines (availability and cost). Of particular concern in Uganda and other low-income countries, is that treatment will be out of reach financially for many people, as newer, more expensive antibiotics will be the only effective treatment options. Additionally, the vast majority of the public in Uganda does not know how antibiotics work or how resistance against antibiotics develops.\(^65\).

One key aspect of antimicrobial resistance in Uganda, Malawi and Kenya is around prescription. It is easy to find antibiotics without any prescription, through community pharmacies. This leads to challenges around the wrong treatment being taken, as well as polypharmacy (the practice of prescribing or taking more than one antibiotic where only one would suffice).

Who?

<table>
<thead>
<tr>
<th>Creators (Participants)</th>
<th>Multidisciplinary teams with expertise in data science and public health (data scientists, clinicians, policy makers, social scientists, behaviour change experts and communicators) could identify new insights and develop new solutions (analysis, products, tools) based on data combined for the first time.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>This could, for instance, take the shape of maps of community pharmacies.</td>
</tr>
<tr>
<td></td>
<td>They could help identify better where community pharmacies are, how antimicrobials flow through the supply chain and where they</td>
</tr>
</tbody>
</table>

are used. This would in turn help to identify “hotspots” for priority intervention and communications.

| Decision makers | • Policy-makers could develop better strategies and guidelines for antibiotic prescription and dispensing, based on a spatial analysis of antibiotics availability in their country. This would help make better decisions on interventions and funding for health policy.  
• Clinicians and healthcare providers could develop better strategies and plans that improve access for people that really need them, while encouraging appropriate use of antibiotics with individual patients, but protect medicines of last resort. Understanding the scale of availability of particular medicines in the surrounding area, they will be able inform patients on the risks.  
• Those advocating for appropriate use of antibiotics with the general public could develop strategies for targeting community interventions and incentivising optimal behaviour. |
| Beneficiaries | • Communities that access antibiotics over the counter who will have a lower risk of developing resistance to antibiotics. |

### Background information

Antimicrobial resistance is one of the major public health concerns of this century. Antibiotics have been a staple of modern medicine’s toolkit to fight illnesses, but antibiotic treatment has become less reliable as antimicrobial resistance (AMR) has evolved in many bacterial pathogens, causing the need for costly new treatments to be developed, while antimicrobial resistant organisms continue to take thousands of lives per year.86

An antimicrobial is any substance of natural, semisynthetic or synthetic origin that kills microorganisms or inhibits their growth. AMR is the ability of microorganisms such as bacteria, viruses and parasites to stop antimicrobials (e.g. antibiotics and antivirals) working against them. This can result in standard treatments becoming ineffective, leading to persistent, difficult-to-treat infections which then may spread to others.

In May 2017, the World Health Organisation (WHO) adopted the global action plan (GAP) on antimicrobial resistance with the goal of ensuring continuity of successful treatment and prevention of infectious diseases.87 In 2019, the WHO followed up with an implementation guide based on learnings from the previous two years.88

---

To help member states craft their own national AMR plans, the WHO partnered with the Food and Agriculture Organization (FAO) and the World Organisation for Animal Health (OIE) to create a manual for the development of a national AMR plan. National action plans (NAPs) are needed in conjunction with the global plan due to the specific needs, circumstances and available resources of each individual country. Kenya, Malawi and Uganda all currently have national AMR plans running until at least 2022.

The WHO suggests that all NAPs adhere to the five strategic goals of the global action plan with contextualisation recommended to ensure local feasibility. The five goals are:

- Improve awareness and understanding of antimicrobial resistance through effective communication, education and training.
- Strengthen the knowledge and evidence base through surveillance and research.
- Reduce the incidence of infection through effective sanitation, hygiene and infection prevention measures.
- Optimise the use of antimicrobial medicines in human and animal health.
- Develop the economic case for sustainable investment that takes account of the needs of all countries, and increase investment in new medicines, diagnostic tools, vaccines and other interventions.

The NAP manual explicitly states that “a robust situational analysis involves assessment and analysis of information on perceptions and behaviour related to known drivers of AMR”. This includes limited knowledge of the risk for AMR; inappropriate use of antibiotics in humans, animals and plants, including overprescribing and dispensing; and incomplete treatment courses (patients who do not finish a full treatment course).

**Kenya**

In June 2017, the Government of Kenya published the “National Action Plan on Prevention and Containment of Antimicrobial Resistance 2017-2022”. The NAP for Kenya adheres to the five strategic objectives of the global action plan. Kenya is also developing a monitoring and evaluation framework including targets to be reviewed halfway through the plan in 2019 and in 2021 before plan completion.

**Uganda**

The Government of Uganda released its national action plan in 2018 to prevent, slow down, and control the spread of resistant organisms while ensuring the continuous availability of safe, effective, efficacious and quality-assured antimicrobials and their optimal use. The NAP also adheres to the five strategic objectives of the global plan.

---

91 ibid
There are important inequalities in terms of access to medication in Uganda. The middle and upper classes get easy access while the rest of the population need to rely on informal ways to get medication. This means a big part of the population only gets access to the medication they can afford (instead of having access to what they actually need).  

It is worth noting that a majority of the health system relies on private institutions. Private institutions might be an opportunity to get access to data more easily - while with public institutions, clearance from the government might be needed.

**Malawi**

Malawi has yet to enact a separate national action plan for its AMR strategy. Malawi developed and implemented a Health Sector Strategic Plan I (HSSPI) from 2011 to 2016, followed by the Health Sector Strategic Plan II (HSSPII), running from 2017 to 2022. The plans are aimed at improvements across the health sector, with AMR included as a critical area of intervention. HSSPI therefore provides a framework for the development of the national AMR strategy in lieu of a dedicated national action plan. According to an analysis produced by the University of Malawi the objectives of Malawi’s AMR plan are in line with the GAP, while an external evaluation by WHO noted that despite having a formal strategy, Malawi was missing some key aspects of the plan such as pharmaceutical production and AMR surveillance for animal health. Though not following the NAP manual exactly, Malawi’s AMR strategy includes introducing and enforcing use of prescription forms at all levels of the health system and strengthening legislation on dispensing and prescription of antimicrobial drugs. In Malawi, a majority of health providers are public, especially in urban places. There are a few private institutions in some rural areas.

**Ghana**

Ghana has been suggested later in the scoping process, as a potential country to look into for this type of challenge. However, further investigation and research would be needed to determine the current context in Ghana. We would need to identify how a data challenge could solve some of the issues related to AMR in the country, what the capacity on the ground is, and what type of data we could rely on.

---

94 Mark Okello interview  
95 Godfrey Omony’s interview  
https://extranet.who.int/countryplanningcycles/sites/default/files/planning_cycle_repository/malawi/health_sector_strategic_plan_ii_030417_smt_dps.pdf  
97 University of Malawi (2019), “Malawi AMR Strategy: some issues to consider during implementation”,  
99 ibid  
100 Malvuto Mndau’s interview
Theory of change

The full theory of change can be found in Appendix 1, below is a broad hypothesis to allow for various streams depending on what makes sense and is feasible in different countries:

- **If we:** run a challenge for data scientists, clinicians, policy makers, social scientists, behaviour change experts and communicators on antimicrobial resistance in Uganda, Malawi and Kenya using antibiotic acquisition and consumption data as well as antimicrobial surveillance data101,
- **Then:** multidisciplinary teams could use insights on antibiotic use, prescribing and dispensing practices and on how these practices drive antimicrobial resistance to develop analyses, tools and products, such as community pharmacies and antimicrobial flow maps,
- **Which:** will help policy-makers, healthcare providers, clinicians and those advocating for appropriate use of antibiotics identify “hotspots” for priority interventions and communications102,
- **Ultimately leading to:** lower incidence of locally untreatable bacterial infections in Uganda, Malawi and Kenya

Likely types of benefit

The challenge prize primarily aims at developing solutions for optimising the current antibiotic use, prescribing and dispensing practices in order to reduce the impact of antimicrobial resistance.

With better analyses, tools and products, policy-makers can develop better strategies and guidelines for antibiotic prescription and dispensing, clinicians and healthcare providers can develop better strategies and plans that improve access for people that really need them, while encouraging appropriate use of antibiotics, but protect medicines of last resort, and those advocating for appropriate use of antibiotics can develop strategies for incentivizing optimal behaviour.

The solutions to the challenge prize might include community pharmacy maps and new analyses of antibiotics use based on the combination of new sources of data. Currently, there is a lack of data related to community pharmacies as they illustrate an informal way of getting antibiotics103. Combining new sources of data and producing such maps would help identify correlations between them and antimicrobial resistance in more details, and would help target interventions.

Policy-makers could develop better strategies and guidelines for antibiotic prescription and dispensing. They could for instance deploy clinicians and pharmacists in certain areas to ensure practices in community pharmacies are in line with national guidelines. They could follow the

---

101 Data that monitors the extent of resistance by different organisms to particular drugs
103 Victoria Rutter & Dr Diane Ashiru-Oredope interview
example in Tanzania with the deployment of accredited drug dispensing outlets (ADDO)\textsuperscript{104} in rural and hard to reach areas, with staff trained to properly dispense antibiotics. Clinicians and healthcare providers such as hospital pharmacists could help inform patients on the risks related to resistance and getting antibiotics via community pharmacies in their area. Those advocating for appropriate use of antibiotics could target and prioritise specific communities or areas for intervention (such as awareness campaigns), and help make behaviours change.

Success indicators

**Impact indicator**: Reduced mortality and morbidity from drug-resistant infections in Uganda, Malawi and Kenya due to more appropriate use of antibiotics.

**Mid-term outcome indicator**: Use of products and tools created during the challenge to help policymakers, clinicians, and advocates build on new data combinations and insights, such as location of community pharmacies and their impact on AMR, to drive the development of better strategies and plans for optimising the use of antibiotics.

The research community advocating for mortality and morbidity data related to drug-resistant infections to be included into the annual Global Burden of Disease Study\textsuperscript{105}. This would be key to help monitor the impact indicator. Using proxy measurements in the absence of reliable data would be an alternative to track this indicator.

To monitor the mid-term outcome indicator, the organisation running the challenge could implement a survey to measure the use of products and tools among the key stakeholder groups identified.

This challenge will meet Wellcome’s intended aims for the challenge prizes related to public trust and capacity building, in the following ways:

- **Increasing public trust**: as all insights and tools will be based on early engagement with relevant communities in the countries, and will be shared widely
- **Increasing capacity**: as training and mentorship will happen throughout the data challenge programme, making sure participants have the capacity to take part and build long term data science capacity that will benefit the country.

Potential time to impact

There could potentially be visible impact within a medium term (two to five years) following the prize.

\textsuperscript{104} ADDO (Accredited drug dispensing outlets) program, https://www.msh.org/journal-tags/addo-accredited-drug-dispensing-outlets-program and https://twitter.com/Wellcome_AMR/status/1062332356277755904

Strategic alignment with Wellcome

Wellcome have an ongoing commitment to transform the world’s approach towards stemming the rise and spread of drug-resistant infections, focusing on the overuse and inappropriate use of antibiotics.

This proposal aligns directly with the Drug Resistant Infections priority area focus on building capacity in Uganda, Malawi, Kenya to collect better quality epidemiological data. This means Wellcome has experience and existing relationships in these countries. It is also in line with initiatives the DRI priority area are closely connected with, such as the Drivers of Resistance in Uganda and Malawi (DRUM) and Global Research on Antimicrobial Resistance (GRAM) initiatives.

This proposal also aligns with Wellcome’s public engagement team’s work on behaviours towards antibiotics use.

Data landscape summary

There are some initiatives currently underway to collect new data and to build capacity in specific countries including Kenya, Uganda and Malawi. A data challenge would need to build on these initiatives and would be dependent on these initiatives to be successful. The amount, quality and access to current data would not be enough to build a data challenge around it.

In order for the challenge to be successful, there would be a need to partner with research projects already on the ground, and coordinate with government surveillance initiatives\textsuperscript{106}. This would allow for better access to data. In the 3 countries, it will also be crucial to rely on data scientists on the ground to be involved both in the data collection and data analysis. This will allow for a better interpretation of the data. Some capacity will need to be built to make sure data analysis is made in the right way, and that decisions can be drawn from it (capacity to interpret the results)\textsuperscript{107}.

AMR surveillance data context

In 2016, following the 2015 publication of WHO’s \textit{Global action plan on antimicrobial resistance}, governments and companies from across the globe committed through two key declarations to tackling AMR: the \textit{Endorsed Political Declaration on AMR} at the United Nations General Assembly (commitment from all UN member states); and the \textit{Declaration on Combating Antimicrobial Resistance} (the Davos Declaration) at the World Economic Forum which included over 80 signatories from pharmaceutical companies.

These declarations included commitments to share AMR surveillance data: data that monitors the extent of resistance by different organisms to particular drugs, as a \textit{key component in the fight against AMR}. Sharing of AMR surveillance data aims to track changes, aid early detection

\textsuperscript{106} Richard Maude interview
\textsuperscript{107} Richard Maude interview
of resistant strains, support timely communications and investigations of outbreaks, inform clinical therapy decisions, guide policy recommendations, and assess the efficacy of existing interventions.

Early 2018, Wellcome and the Open Data Institute worked on a pilot project putting together an AMR register, focused on gathering and openly publishing ‘raw’ data on industry sponsored human antimicrobial surveillance programmes. Wellcome organised a data re-use making available surveillance datasets from the ATLAS-Pfizer study, in November 2018. There are ongoing conversations with other pharmaceutical companies, such as Shionogi and GSKto potentially add more data to the register (the AMR benchmark is pushing companies to engage more).

The AMR register data provides an overview of age range, country and time, inpatient vs outpatient, type of infection, etc. It does not contain any epidemiological information which could inform on the health status of the population. Additionally It does not provide information about patients outcomes (the data doesn’t show what antibiotic was prescribed to the patient or what the outcome of their treatment was), as this information is not routinely collected by pharma companies. There is a need to streamline data collection, analysis and sharing to better inform treatment of patients with resistant infections. Evidence generated will also inform policy-makers in developing and implementing public health interventions that respond to local/regional and country needs.

Challenges around current data sources

Data on community pharmacies\textsuperscript{108}
In various countries in Africa, including Uganda, Malawi and Kenya, regulations on medicine sellers vary. As a lot of antibiotics are provided through community pharmacies that are unregulated, it makes it hard to collect data on antibiotics stock and use in an area. In some communities, there is no data related to which antibiotics people are taking, how much, and where from.

Point prevalence surveys
Point prevalence surveys collect information on prescribing practices of antibiotics and other information relevant to treatment and management of infectious diseases in hospitalized patients, and complements surveillance of antimicrobial consumption\textsuperscript{109}. Even though such surveys could serve as a good basis for a challenge, some data scientists consider them as collecting data that is not good enough for data analysis\textsuperscript{110}.

Routine health facility surveys\textsuperscript{111}
Routine health facility surveys provide a valuable data source on the availability and use of antibiotics in low- and middle-income countries.

\textsuperscript{108} CPA interview
\textsuperscript{110} Tim Jinks interview
\textsuperscript{111} Rebecca Knowles, Mike Sharland, Yingfen Hsia, Nicola Magrini, Lorenzo Moja, Amani Siyam & Elizabeth Tayler (2019), Measuring antibiotic availability and use in 20 low- and middle-income Countries, https://www.who.int/bulletin/volumes/98/3/19-241349.pdf?ua=1
Antibiotics stocks
Doctors usually record what prescriptions they give. So working with the health facilities could work, and with pharmacies in these facilities (even in terms of stocks, difference between beginning and end of the month).
Doctors usually record what prescriptions they give. Working with the health facilities could be helpful to access some data, and with pharmacies in these facilities (even in terms of stocks, difference between beginning and end of the month). There might be some difficulties accessing data from doctors though. An approach could be to get a single hospital and retrospectively see how many antibiotics were prescribed and try and see if there would be any linkage between the samples and the prescriptions.

Data sources
Available data
- Pfizer Atlas: Pfizer conducts one of the largest global surveillance programs in the world, called Antimicrobial Testing Leadership and Surveillance (ATLAS). ATLAS features 14 years of cumulative data on changing bacterial resistance patterns across more than 60 countries.
- Antibiotic Resistance: Interdisciplinary Action (AR:IA): application developed through Wellcome’s Data reuse prize
- IHME - Global Burden of Disease (GBD): the Global Burden of Disease (GBD) data are freely available. It contains global, regional, national, and in some cases subnational estimates of the burden of diseases, injuries, and risk factors. Visualisations available too.
- WHO - GLASS: the Global Antimicrobial Resistance Surveillance System (GLASS) promotes and supports a standardized approach to the collection, analysis and sharing of AMR data at a global level by encouraging and facilitating the establishment of national AMR surveillance systems that are capable of monitoring AMR trends and producing reliable and comparable data.
- IQVIA: IQVIA is committed to identifying and understanding the driving factors behind the rising number of drug-resistant bacteria strains to inform ongoing prevention efforts. To improve antibiotic use, IQVIA is committed to building on an existing framework of antibiotic prescribing reporting and analysis. Access to this data is limited and needs to be paid for.

Data in the process of being collected
The AMR register, a pilot project, ran in 2018, focused on gathering and openly publishing data on industry sponsored human antimicrobial surveillance programmes. Wellcome’s DRI team is currently working with a pharmaceutical company (to be announced in the summer 2020) to strengthen AMR surveillance in Ghana, Kenya, Malawi and Uganda The initiative aims at collecting and publishing epidemiological and susceptibility data. This will include 20 isolates for each bug for 5 pathogens in 4 countries including Ghana, Kenya, Malawi and Uganda across 200 patients per year in one facility in each country. In each country, up to 2 sites/ laboratories will be recruited. The idea is to build capacity in these countries to collect data (to be collected by the end of 2020 or early 2021).

Examples of data to be collected will be on isolates and on patients.
   1) Isolate / subject number

112 Dr Sangoro interview
- Isolate eligibility
- Isolate collection day (Note: this will not be a date, but will be designated as day 0 to support elements below)
- Isolate identification
- Isolate susceptibility testing results (site laboratory DST – drug susceptibility testing)

2) Patient data (not published on the AMR register as it will contain sensitive information, accessible through a clinical trial platform)
   - Demography - age, sex
   - Presence of comorbidities
   - Presence/absence of clinical signs and symptoms of infection
   - Was one or more of the following signs and symptoms of systemic inflammation present at the time of specimen collection?
     - Fever ≥ 37.5°C
     - White blood cell (WBC) count of >10,000 cells/mm3
     - Local/regional lymphadenopathy
     - Elevated C-reactive protein (CRP) (nonspecific)
   - Site of Infection
   - Infection treatment
   - Hospitalisation status
   - Mortality within the protocol window.

Wellcome’s Public engagement team is scoping some work around the collection of citizen generated data related to DRI. The idea would be to collect data on antibiotic acquisition and consumption, especially related to children (through a mix of qualitative and quantitative methods).

Data is also currently being collected as part of the GRAM\textsuperscript{113} (Global Research on Antimicrobial Resistance) and DRUM\textsuperscript{114} (Drivers of Resistance in Uganda and Malawi) projects.
With GRAM, maps demonstrating AMR burden levels globally will be published at the end of June 2020. The goal is to include figures in IHME’s Global Burden of Disease Study.
With DRUM, wrapping up in April 2021, some a key source of data would be around behaviours related to antibiotic use in Uganda and Malawi.

\textsuperscript{113} https://www.bdi.ox.ac.uk/oxfordgbgroup/research
\textsuperscript{114} https://qtr.ukri.org/projects/ref=MP%2FS004793%2F1
Challenge structure

The challenge period will cover a period of 18 months, with a 9 month set up phase preceding the prize, and 9 month sustainability prize following the judging.

For this topic, due to the Covid-19 context, there will be a need to reassess the focus and the model. Consequences of Covid-19 on antibiotic consumption, and on capacity on the ground in Africa, could be considerable. A data challenge on DRI in Africa could only be put together once we have a better idea of the consequences of Covid-19.

The hypothesis / challenge focus could be adapted to each country. Depending on the situation related to data availability and policy priorities, the challenge could encourage new data analysis for different purposes, such as better decision making or more efficient public engagement interventions.

The challenge could thus be run following different models: one specific challenge per country, or one broad challenge across the 3 countries.

Seed funding would be needed in order to allow for people to form multi-disciplinary teams and take part in the competition. Details of this are available in the budget and in the descriptions below.

Training and mentoring would also be required, to align capacity and to ensure benefits would be lasting for the participants and the country as a whole. Specific training requirements will be established early in the challenge, depending on what participant needs are. Training might cover relevant areas from the data skills framework, such as ethical use of data or machine learning approaches.

This challenge could be run with the following partners:
- Africa Center for Disease Control (CDC) for expertise and connexions with ministries of health
- Commonwealth Pharmacists Association (CPA) for expertise and connexions with relevant stakeholders in the healthcare systems nationally and internationally
- Tropical Health And Education Trust (THET) for expertise and connexions.
- Center for Disease Dynamics, Economics and Policy (CDDEP) for expertise on data analysis.
- GRAM project team for access to maps showing the burden of antimicrobial resistance worldwide
- DRUM consortium for access to data as well as access to data scientist communities in the target countries

Locations and participant recruitment

ODI’s recommendation is that while the three countries present many opportunities, the challenge should be delivered in a single country, not all simultaneously. Therefore Wellcome should use the information in this report along with further consultation with stakeholders to prioritise one country over another (Kenya, Malawi, Uganda). It would be possible to run the
same challenge sequentially first in one country, and then the two others. This has not been accounted for in the recommended sequencing or budget, however it is feasible and if the initial challenge is successful it would provide a solid framework for repeating in other locations.

In terms of recruitment for participants, there are extensive domestic data science networks across these regions, such as Data Science Africa and Machine Intelligence Africa in addition to country specific networks. The recommendation therefore is to focus on these networks for recruitment. This will have the consequence of building on the existing links across the networks to enable a pan-African support network.

A specific set of end user beneficiaries and community representatives will need to be engaged throughout the design and delivery of the challenge prize. Participants should be recruited from across the country, and supported to travel to key events. This requirement for expenses is reflected in the budget. This is intended to enable a peer network of local experts across the country.

Remit and scope of the challenge

Key question
How can we use data to optimise the use of antibiotics in order to minimise the local emergence and spread of antimicrobial resistance in Uganda, Malawi and Kenya?

Suitable Proposals

Proposals should be submitted to make progress in one or more of the following areas:

- **Insights or models** that can provide insights about antibiotic acquisition and consumption patterns, prescribing and dispensing practices and on how these drive antimicrobial resistance
- **Tools/ applications** that can inform health policy-makers around **policy priorities**, for example using mapping across the country to identify where deployment of professionals or more formal pharmacies would help
- **Tools/ applications** that can be used by healthcare professionals to provide more **informed advice on antibiotic prescribing and consumption**
- **Tools/ applications** which allow advocates to target the right areas for **awareness campaigns and strategies** to change behaviours towards antibiotic consumption.

Proposals can be submitted to provide solutions in a single selected country or across multiple countries.

Any proposal submitted should consider the end goal of enabling policy-makers, healthcare providers, clinicians and those advocating for appropriate use of antibiotics to develop strategies for optimising the use of antibiotics. The proposals should consider the particular country context and existing antibiotic use, prescribing and dispensing practices.
Who can apply

We are looking for multidisciplinary teams who bring data science together with a range of alternative disciplines in order to deliver innovative solutions to this complex issue. Example disciplines include:

- public health
- computer science and machine learning
- digital development
- service design

The prize will be open to all international participants, but there will be significant engagement and recruiting with local networks of experts and all events will be run in the location specific to the prize.

The application process is open to those who have predefined ideas and projects, and those who may have relevant skills but no project proposal.

Participant motivations

The motivation for participants to enter will be a combination of:

- opportunities for networking / building relationships
  - access to key government and healthcare sector stakeholders
  - access to other experts within and across disciplines
- personal and professional development
  - advancing concepts and technologies the field of antimicrobial resistance research
  - advancing individual capabilities
- personal and professional rewards
  - financial reward of winning the prize
  - reputational reward of winning the prize

The motivation for academic participants tends to be more driven by access to key government and healthcare stakeholders, improvement in and contributions to their field of study, and reputational awards.

The motivation for commercial participants tends to be more driven by access to new markets and clients, positive press around service provision and financial rewards.

The motivation for governmental stakeholders will be to outsource the development of innovations that can help them deliver more efficient services, and contribute towards SDG goals.
Challenge stages, funding & support

The estimated time allocations for the DRI data challenge would include:

- **Up to 9 months** for a set up phase
- **Up to 18 months** for the challenge to run, from recruitment of participants through to the awarding of the challenge prizes
- **Up to an additional 9 months** for a sustainability period

<table>
<thead>
<tr>
<th>Reflection &amp; Evaluation</th>
<th>Set up and recruitment</th>
<th>Launch &amp; Applications</th>
<th>Insights phase</th>
<th>Development phase</th>
<th>Judging Period</th>
<th>Sustainability Period</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

For a detailed view of estimated timelines, please refer to the DRI data challenge timeline in Appendix 10.

1. **Set up phase**
   The organisation that is selected to run this challenge programme would likely need six to nine months in order to set up the challenge programme. This might include, but is not limited to, the following activities:
   1. Convening beneficiaries and stakeholders across the communities in question, local, state and national government and private sector in order to verify their needs with regard to antimicrobial resistance and confirm their engagement with the prize.
   2. Establish agreements with local delivery partners, and establish the detailed delivery plan.
   3. Establish agreements with data providers.
   4. All data will need to be vetted and cleaned thoroughly prior to launch in order to verify quality and limitations. This will take dedicated resources and be the most time consuming part. The aim is to ensure that participants can use the data immediately at the point they enter the Insights Phase.

2. **Participant recruitment/ Expression of interest**
   A period of up to two months will be required for recruitment of participants, and this can happen concurrently with the end of the set up period. A website will need to be established with all the required information on topic and processes. This will then be publicised through some advertising and media channels. In this case, it is likely that the bulk of the participants will be recruited through established networks of data scientists, or through existing networks such as the Commonwealth Pharmacists Association.
Participants will be invited to complete an expression of interest (EOI). EOIs can be open to those who have predefined ideas and projects, and those who may have relevant skills but no project proposal. Participants will be able to communicate with each other during this period, and will be encouraged to make connections and discuss ideas to help with team formation.

3. **Launch lab**
The challenge prize will begin with a ‘lab’ or multi-day event (possibly virtual given the Covid-19 crisis). All those who have completed an EOI will be invited, along with beneficiaries, data providers and delivery partners. Participants will have their expenses paid for travel and accommodation.

The first part of this event will focus on explaining why this challenge topic is so important, telling the stories of why and how this problem needs to be addressed, and the processes involved in participating in the challenge prize. This can be recorded to be sent to people who cannot attend. Press will be invited to cover the ‘launch’ and meet some of the participants and key stakeholders.

Following this introductory session, participants will have the opportunity to share their project ideas, recruit new team members, and engage with data providers and stakeholders.

The aim is to support the development of multidisciplinary teams, bring them into contact with the data providers and beneficiaries, and kick start the development of their proposals and ideas through two days of workshops and activities. There will be a clear process for how to team up and formally enter as a team.

4. **Team applications**
The teams will be given 4-8 weeks following the launch lab to refine their proposal ideas, confirm their teams and submit their area of interest/initial proposal in a formal team application. This will then be assessed based on criteria including:

- Multidisciplinary team with relevant expertise
- Area of interest in line with challenge scope
- Scale of ambition

During this period, participants will have access to a challenge team application template to complete, have access to supporting information on the topic, and will be able to troubleshoot with the organisation running the challenge for process queries.

Teams may submit their proposal at any time during this period and begin the next phase, though there will not be access to funding before 4 weeks. The application window will remain open until 8 weeks after the launch lab.

5. **Model development phase**
This phase will be between 6-9 months and focus on training and the development of data insights. The teams will be encouraged to focus on engaging with the available data and modeling it to develop insights. During this time they will receive some seed funding (£25k for up to 10 teams), data access support to help them identify the data
they need for their project and any governance requirements, and a rolling series of seminars on relevant machine learning techniques.

They will also be receiving some initial training and access to mentors, to ensure they are provided with opportunities to develop these insights.

In this phase we would expect the members of the teams with data science and public health background to be leading the work.

Through the development of models and insights, the teams will be asked to put together plans for developing practical applications and tools for end-users drawing on the data science methods and insights they have developed. These tools and applications will need to address specific needs of beneficiary groups such as, but not limited to:

a. Insights or models that can provide insights about antibiotic acquisition and consumption patterns, prescribing and dispensing practices and on how these drive antimicrobial resistance
b. Tools/ applications that can inform health policy-makers around policy priorities, for example using mapping across the country to identify where deployment of professionals or more formal pharmacies would help
c. Tools/ applications that can be used by healthcare professionals to provide more informed advice on antibiotic prescribing and consumption
d. Tools/ applications which allow advocates to target the right areas for awareness campaigns and strategies to change behaviours towards antibiotic consumption.

The model development phase will culminate in a demo day in which all of the teams will demo the models and insights they have developed using the data, and their pitches for tool development. These will be judged by a panel of key stakeholders and experts in the field against a set publicly accessible scoring criteria. For example, applications might be assessed on some of the following areas:

- Understanding of the problem
- The proposed approach
- Team resources
- Proposed outcomes
- Potential impacts
- Project management documentation
- Risk assessment
- Cost plan
- Ethical assessment

The best teams will be selected to go through to the development stage to build their proposed tools/ applications. The number of successful applicants who move from the Insights Phase to the Development Phase would likely be pre-agreed at this stage, in order to control costs in the discovery phase. Alternatively applications could be judged on merit, but this would mean there is no limit to the number of participants in the first phase and would cost more to run.

6. **Tool development phase**
The tool development phase will also run for between 6-9 months. The projects that have progressed will receive further seed funding (£100k for up to 5 teams) in order to develop the tool or application outlined in the insights phase.

In this phase we would expect the teams to have a greater focus on digital development, working with and drawing on the expertise of the team’s scientific experts and stakeholders to inform their work. These teams will receive additional training and mentorship suitable to their specific topic to enable their ideas to develop and flourish. This will include support to plan implementation with stakeholders, ethical assessments and business model development.

Where possible, the challenge teams will get a chance to implement their full or partial solutions in an environment with real stakeholders. This will allow for feedback, adjustments and further development before the final judging. This will need to have been agreed in principle with key stakeholders during the implementation phase, and then coordinated by the delivery team following the insights period.

7. Judging
The development phases will culminate in a public demo day attended by as many media representatives and senior stakeholders attending as possible. Each of the teams present their work. The event will have a strong focus on storytelling - what the nature and scale of the risks around antimicrobial resistance are, and how we have used data science to address them.

A set of judges who are known experts in the field, such as Rob Terry (WHO), Karen Bett (Data4SDGs Kenya), Diane Ashiru Oredope (PHE / CPA), Catrin Moore (Oxford Big Data Institute, GRAM), Nick Feasy (DRUM), Chris Jewell (Lancaster University), Marc Henrion (Malawi Liverpool Wellcome Trust), will review the submissions against publicly available criteria. Criteria are likely to include:

- Innovation
  - Ethical data management
  - Equitable solution
  - Utilises multiple data sources
- Impact
  - Scale of potential impact on antimicrobial resistance in the region
  - Clear plans on how the solution could improve public trust
- Capacity
  - Clear input from multiple disciplines
- Beneficiaries
  - Clear link to government/ other appropriate end user
  - Scalable solution which has the potential to be used in other regions
- Sustainability
  - A clear, time bound implementation plan for the solutions to be adopted by end users
  - Whether the solution have the potential to continue beyond the prize

The most successful submission will be awarded a prize of £200k. Three runner up teams shall receive £100k. This money will be explicitly for the implementation and scaling of the tool they have developed. The rationale for this model is that it allows multiple tools and outputs to go on to be scaled and supported, delivering sustainable
impact in the field. This has been benchmarked\textsuperscript{115} against other health technology innovation prizes.

8. **Sustainability/ Uptake period**
   It is vital that there is a sustainability period following the award of the prize in order to support the winning project and the other projects to be adopted by the key stakeholders and used to address the key issues addressed regarding mental health. The winning teams can use the prize money to implement and scale up the innovations they have developed.

   In addition, the partners involved in the delivery of the prize will need to remain available for the teams over a period of 9 months in order to:
   - Assist the future development of winning solutions by making introductions to would-be users, investors, commissioners and others.
   - Establish a schedule for checking progress with winners and interesting other participants.
   - Seek feedback from competitors, judges and partners and evaluate the effectiveness of the programme.
   - Return to the prize winners a year afterwards to develop a case study of where they have got to in their development and the impact they have created.

9. **Evaluation**
   At the end of the judging and sustainability phase there shall be an independent evaluation made of the prize as a whole drawing on reporting material and interviews with participants and stakeholders. The interviews will also be used to develop case studies of each output and prize in order to share learnings and successes for others interested in the field of health data innovation and remote sensing for public health.

**Risks**

A full risk register is available in Appendix 3. Key risks for this data challenge are related to 3 main topics:

- This challenge is relying on sources of data coming from current or future initiatives (such as the collection of behavioral data or better quality epidemiological data in the 3 countries). We would need these initiatives to be successful and to grant access to data, for this challenge to be successful.
- Across the 3 countries, in major hospitals and health centers, data can sometimes be collected electronically. However, there are some limitations to the availability of data (some communities do not have access to healthcare, only parts of the patient journey are recorded, data can mostly be found in urban areas while most of the population lives in rural areas)
- the changing context worldwide, related to Covid-19 and any upcoming health crisis. Access and consumption of antibiotics can be completely altered in a situation such as Covid-19, with people attempting various treatments to try and stop or cure it. Priorities related to drug-resistant infections will need to be reviewed in light of this context.

\textsuperscript{115} Appendix 9
Challenge prize data ecosystem

An ecosystem map can be helpful in understanding how data flows between different stakeholders involved in snakebite management and how value is created in the ecosystem by allowing different stakeholders to make better decisions.

This ecosystem map illustrates how several existing initiatives, producing key data related to antibiotic consumptions and antimicrobial resistance, can drive challenge insights (creating spatial overview, helping identify priority areas for interventions). These types of insights, created through the combination of new data being collected, can have consequences on the health sector as a whole: through policy makers, health professionals and activists.

**DRI Challenge Data Ecosystem Map.** A high res pdf of the map is available in Appendix 13 and an interactive version is available [here](#).

![Ecosystem Map](image)

In the 3 countries, one of the key next steps to put together a data challenge will be to engage with Ministries of health, to explain what we are trying to achieve and demonstrate the interest for public health in their country. It would be necessary to get their buy-in early on.

In order to gain trust and buy-in from key stakeholders in Uganda, there will be a need to explain this data challenge would not be a “one-off”, and that the results would benefit the health sector and the country as a whole\(^\text{116}\).

---

\(^{116}\) Godfrey Omony interview
## Capacity and stakeholder landscape

Across the 3 countries, in major hospitals and health centers, data can sometimes be collected electronically. However there are some limitations to the availability of data:

- Some communities do not have access to health care (data might be missing for marginalized communities)
- Health problems have been recorded, but only the diagnosis and treatment are captured (other data might be unstructured and incomplete)
- Most of the health facilities with electronic data capture are in the urban setting yet a majority of the population lives in the rural areas.

Many stakeholders are involved in antimicrobial resistance, they can come from various perspectives such as policy, advocacy, funding, innovation and research, as well as surveillance. This is the list of key organisations to be considered and consulted with in delivering this challenge.

| Research / Academia / Activism | Tropical Health And Education Trust (THET)  
|                               | Commonwealth Pharmacists association (Potential partner: they are already partnering with specific hospitals to collect data (could be an entry point to specific hospitals as data stewards / specific places to run the data challenge)  
|                               | Africa CDC (Potential partner)  
|                               | Zindi Africa (Potential partner)  
|                               | Kenya Medical Research Institute (KEMRI)-Wellcome  
|                               | African Academy of Sciences (potential partner)  
| Technology Institutes          | Data Scientists in Kenya, Malawi, Uganda (contacted through Shawn Dolley's work, all interested in taking part in the challenge and engage with their community)  
| Healthcare                      | For each country, medical universities  
| Co-Funders                      | Wellcome DRI team + Team awarded / helping them with capacity building in Uganda, Malawi, Kenya  
|                                | Wellcome public engagement team + Team awarded to run the citizens generated data project  
|                                | Fleming fund  
|                                | USAID  
| International bodies           | World Health Organization (WHO, through TDR)  
|                                | United Nations (especially Data4SDGs) - Potential partner: they’re building capacity / working closely with the Ministry of Health on interoperability

Govt and Regulators |  ● For each country:
| | ○ Ministry of Health - Necessary to get their buy-in if we want to run data challenges for their country
| | ○ Council of Medical Research / Medical Council
| | ○ Data Protection Authority

Individuals |  ● Potential data stewards if we take the route of citizen generated data

Data science skills

**Uganda** In Kampala, there is an important data scientists community. However there is still a need to train people in order to get them to participate in a challenge\(^{118}\).

There might be some specific parts of the country with more capacity than others. It is the case in the North of the country for instance, where a Data Science Center was established in June 2019 and where some hospitals such as the one in Gulu might have more data to share\(^{119}\).

**Kenya** is a country with a strong data community\(^{120}\). The Kenyan community is quite vibrant, there could be many ways to engage (tv shows, radio shows, sms, etc). There would not be a “one size fits all” solution. Young people use phones, elderly would prefer the radio\(^{121}\).

**Malawi** There is a growing data science community in Malawi. Some universities, such as the Mzuzu University (second largest in Malawi) have active ICT departments\(^{122}\).

Healthcare data providers

**Uganda**

There is an opportunity to work in Uganda, given the quality of the lab capacity. Engaging with experts in Uganda, Wellcome teams observed how Uganda wanted to actively be more competitive internationally. Some training could be included in the design of the programme.

In terms of access to data, a majority of health services are provided by the private sector, so discussions would need to happen with particular companies. For public institutions such as hospitals, the Ministry of Health would need to be included in conversations. A key challenge in Uganda is that a lot of data is on paper. As there are very poor records, it is hard to get access to the medication history of a patient (making it even harder to fight against AMR)\(^{123}\).

However, in the past 2 or 3 years, Uganda has shown some signs of progress towards more data infrastructure. Some centers for data science at district and national level were created, and initiatives such as the [Raxio data center](#) were launched.

---

\(^{118}\) [Mark Okello interview](#)
\(^{119}\) [Godfrey Omony interview](#)
\(^{120}\) [Data4SDGs interview](#)
\(^{121}\) [Dr Sangoro interview](#)
\(^{122}\) [Chimango Nyasulu interview](#)
\(^{123}\) [Esau Lwanga interview](#)
Including some training and mentorship into the data challenge would not only be beneficial to data scientists but also draw buy-in (people would trust the data more)124.

**Malawi**125
There are no electronic medical records yet. Registers in health facilities would be a good place to look at, however they are on paper126.
There are issues related to various languages being present in Malawi (making it hard to work with consistent data).
There are also challenges related to a lack of access to good equipment such as powerful computers, fast internet (being costly). There are also few experts in the field of health data.
This means there would be some training and seed funding needed to run a challenge in Malawi, in order to motivate new interests.

**Kenya**
Several initiatives from the Wellcome, Data4SDGs, CPA teams are currently building capacity in Kenya to collect better quality data and to make it more interoperable.

In Kenya, health is managed at a subnational level. This creates some inequalities between counties, some being still catching up on digitisation. Electronic health records (EHR) are now used in a small number of counties (some topics are more digitised than others, such as HIV and Aids). USAID is funding an initiative to help counties across Kenya to transition to digital, some significant progress should happen by the end of 2020.

Kenya is considered as quite an advanced country when it comes to data management. However in the health sector, Kenya is still developing and there is a need for policy changes regarding sharing of information that is highly regulated. Further development of infrastructure for capturing health information is needed as well127.

**Wider challenge landscape**

**Antimicrobial Stewardship programme**128
The Commonwealth Pharmacists Association is leading the The Commonwealth Partnerships for Antimicrobial Stewardship programme. It is funded by the Fleming Fund (an initiative from DFID, providing grants to improve lab capacity and AMR surveillance at a country-level) and is building capacity in countries such as Ghana, Tanzania, Uganda and Zambia to collect AMR data.
For this, some health institutions in the UK are partnering directly with institutions in the target countries. The idea is for pharmacists on the ground to be empowered to conduct data collection.

This will take place through 12 partnerships, with specific hospitals (considered the best way to get data). The partnerships will particularly focus on promoting the rational use of antimicrobials, looking at how they can build capacity for informed antimicrobial prescribing

---

124 Godfrey Omony interview
125 Open Global Health (2019), Southern & East Africa a Landscape Analysis Data for Science and Health
126 Mavuto Mndau interview
127 Open Global Health (2019), Southern & East Africa a Landscape Analysis Data for Science and Health
128 THET + CPA + CDC interview
using appropriate protocols, antimicrobial stewardship and surveillance of use – including through participation in the Global Point Prevalence Survey (Global PPS)\textsuperscript{129}.

In the 4 countries, more point prevalence studies have been done. Through PPS, data is being collected across the hospital, it is used to develop interventions (collecting data, putting an intervention in and reassessing, either for hospital level or patient level).

This programme was initially planned to wrap up in April 2020 but might be delayed due to Covid-19.

**Mapping Antimicrobial Resistance and Antimicrobial Use Partnership (MAAP)\textsuperscript{130}**
The Fleming fund is also aiming at expanding the volume of historical and current data on Antimicrobial Resistance and Antimicrobial Use (AMU) across Africa and Asia, through a mapping initiative.

Mapping Antimicrobial Resistance and Antimicrobial Use Partnership (MAAP) is a consortium of partners responsible for the implementation of the Fleming Fund Regional Grant Program in Africa. The grant was awarded to the consortium in December 2018, for a duration of up to 2 years. The consortium is comprised of seven partners namely African Society for Laboratory Medicine, the Center for Disease Dynamics, Economics and Policy (CDDEP), IQVIA, Africa Centers for Disease Control and Prevention, West African Health Organization, the East Central & Southern Africa Health Community (ECSA-HC), and Innovative Support to Emergencies, Diseases and Disasters (InSTEDD). Researchers are involved in executing the data collection and analysis activities. Data visualization will be done via ResistanceMap (ResMap), and the Drug Resistance Index (DRI), a metric that aggregates antibiotic use and resistance into a single measure, will be used to compare antibiotic effectiveness.

**Global Research on AntiMicrobial resistance (GRAM)**
Global Research on AntiMicrobial resistance (GRAM) is the flagship project of the Oxford GBD (Global Burden of Disease) Group\textsuperscript{131}, and aims to provide robust, comprehensive and timely evidence of the burden of antimicrobial resistance (AMR) globally, and in 195 countries and territories. This project is funded by the UK Department of Health’s Fleming Fund, the Wellcome Trust and the Bill and Melinda Gates Foundation. It is wrapping up in June 2020.

The team is focusing on consolidating, reviewing and analysing all available data and scientific information on AMR worldwide (literature, surveillance systems, vital statistics, etc). They arecombining various sources of data that have never been combined before (such as microbiology data, mortality data, patient data)\textsuperscript{132}. The goal is then to produce granular geospatial maps of AMR burden to enable policymakers and researchers to tailor interventions to the local level, and engage with the wider public through tools and visualisations. AMR will then be included in the Global burden of disease study\textsuperscript{133}. Maps will be published however granular data access is not part of the current grant structure.

---

\textsuperscript{129} About CwPAMS, https://commonwealthpharmacy.org/commonwealth-partnerships-for-antimicrobial-stewardship/

\textsuperscript{130} https://cddep.org/projects/mapping-antimicrobial-resistance-and-antimicrobial-use-partnership-maap/

\textsuperscript{131} The Oxford GBD Group, https://www.bdi.ox.ac.uk/oxfordgbdgrou

\textsuperscript{132} Catrin Moore interview

\textsuperscript{133} http://www.healthdata.org/GBD
Drivers of Resistance in Uganda and Malawi (DRUM)\textsuperscript{134}

The DRUM consortium, led by the Tropical School of Tropical Medicine, aims to build an agent based model of the drivers of antimicrobial resistance transmission, using data from urban, peri-urban and rural settings in Uganda and Malawi to inform the model. This project is planned to wrap-up in April 2021 (likely to be postponed due to Covid-19).

One aspect will focus on collecting behavioural data around why people use antibiotics in Uganda and Malawi, through interviews.

Public engagement on DRI\textsuperscript{135}

The Wellcome public engagement team is involved in initiatives related to DRI in sub-saharan countries. One initiative is around responsive dialogues in Malawi. It aims at creating citizen assemblies types of dialogues, between policy makers, researchers and the public. The idea is to co-create local solutions on the topic of DRI, in the context of national action plans being sometimes difficult to implement.

A second initiative (still in pre-application phase), focuses on collecting citizen generated data in relation to DRI. The idea would be to collect data on antibiotic acquisition and consumption, especially related to children (through a mix of qualitative and quantitative methods).

ESRC funded research project in South Africa\textsuperscript{136}

Research is currently being done by the Centre for Health Policy, University of the Witwatersrand, Johannesburg in South Africa around health workers’ behaviour related to AMR. This research is funded by the Economic and Social Research Council (ESRC), and is delivered in collaboration with the London School of Economics.

In South Africa, access to antibiotics is well regulated, most people get access to antibiotics only via prescriptions, and there is a lot of available data on resistance patterns. Challenges lie in complex behaviour from skilled-healthcare professionals.

For the purpose of this research, they are using standardised patients to find out what is being prescribed during consultations. When testing with a “well-informed” patient, healthcare professionals would prescribe significantly less antibiotics. This proved that public knowledge and public engagement helps. However, there was still a lot of negative behaviour (antibiotics being prescribed to people who did not need it).

There is also an issue of trust in South Africa around the guidelines provided around antibiotics prescriptions. It seems like research and articles published in science journals have more influence than advisory from a government level.

WorldWide Antimalarial Resistance Network (WWARN)\textsuperscript{137}

The WorldWide Antimalarial Resistance Network (WWARN) was created in 2010. Over the past decade, they have navigated the transition from “data sharing [being] a swear word”\textsuperscript{138} to

\textsuperscript{134} Drivers of Resistance in Uganda and Malawi: The DRUM Consortium, https://gtr.ukri.org/projects?ref=MR%2FS004793%2F1
\textsuperscript{135} Wellcome public engagement interview
\textsuperscript{136} Duane Blaauw’s interview
\textsuperscript{137} WWARN case study
\textsuperscript{138} Elizabeth Pisani, Stella Botchway (2016), Learning from the pioneers: lessons about data platforms drawn from the WWARN experience
having gathered, curated and enabled access to trial data from over 180,000 patients. By allowing researchers to access this data, WWARN has made a significant contribution to the fight against drug resistance for one of the deadliest infectious diseases affecting much of the world’s poorest.

WWARN is a network for researchers to share clinical trial, patient level, data. Understanding when, how and where resistance occurs is enhanced through meta-analysis (taking and combining data from lots of different sources to produce new insights) of patient data. IDDO is building on the success of WWARN: a collaborative platform generating innovative resources and reliable evidence to inform the malaria community on the factors affecting the efficacy of antimalarial medicines.¹³⁹

**Special Programme for Research and Training in Tropical Diseases**¹⁴⁰

TDR, the Special Programme for Research and Training in Tropical Diseases, is a global programme of scientific collaboration that helps facilitate, support and influence efforts to combat diseases of poverty. It is hosted by the World Health Organization (WHO). It helps with scoping research and maximising its use to influence decision making.

One of its core activities is around providing structured operational training programmes to health workers. They are trained to use and analyse that data, publish data in academic papers. On the topic of AMR, they worked on an initiative funded by the UK Department of Health and Social Care, in various countries including Uganda, Rwanda, Myanmar, Nepal. Participants are provided with mentors, helping them develop a research protocol based on specific data, training them on data analysis and providing them with 6 months to produce a research paper.

They are also working on Ebola through their Ebola data scientist fellowship model. For this, they are also using IDDO which aggregates and standardises clinical, epidemiological and laboratory data on Ebola infection. They encourage researchers to use this engage with this data.

**MORU Tropical Health Network**¹⁴¹

One of MORU's initiatives is to use data science to drive better decision making at government level, related to malaria. They are deploying data scientists in various countries such as Laos, Bangladesh, Myanmar to collect resistance data and to map it. Data covers blood tests as well as demographic and travel data (looking at spreading trends).

The aim is to present the data back to the ministries of health, in a standardised format, and to accompany them in producing policy recommendations and action plans based on that data. A workshop, a report and the raw data is provided to the ministries (through a closed platform). In the longer term, the idea is to train ministries on how to use the data.

**Health data interoperability**¹⁴²

The Data4SDGs team is currently working with the Ministry of Health in Kenya on data interoperability. This is following the World Data Forum in 2018 following which a guide for

---

¹³⁹ IDDO, Our work, [https://www.iddo.org/about-us/our-work](https://www.iddo.org/about-us/our-work)
¹⁴⁰ [Rob Terry’s interview](https://www.iddo.org/about-us/our-work)
¹⁴¹ Richard Maude interview
¹⁴² [Data4SDGs interview](https://www.iddo.org/about-us/our-work)
governments on data interoperability was published. The team is helping the ministry put the guide into practice, helping data flow more easily between different parts of the government, as well as helping with digitisation (transition to electronic medical records).

This project was set to end in June 2020. However, due to Covid-19 the timeline might be reviewed.

Legal, policy and regulatory environment

Kenya

The Kenyan government has driven the growth of a significant data economy across both public and private sectors as part of its goal of hosting a globally competitive knowledge based economy by the year 2030.\(^{143}\) In 2011, Kenya launched the Kenya Open Data Initiative, the first such programme in Sub-Saharan Africa, making key government data, including in health, freely available through a single open data portal, [www.opendata.go.ke](http://www.opendata.go.ke). Further activities include the launch in 2016 of the Kenya Health Data Collaborative, an initiative with global health partners to improve the availability, quality and use of data for decision-making. In this context, the health data challenge prizes align with the government’s commitment to enabling the innovative use of health data.

In 2019, Kenya passed its Data Protection Act implementing Article 31 of the Constitution of Kenya (2010), the right to privacy. The Act harmonized the legal framework on data protection, becoming the primary and comprehensive statute. The Act defines key terms and roles, and makes clear requirements for express and informed consent for the collection and processing of personal data. The provisions in the Act are similar to those applying to international standards (particularly GDPR) which should ensure ease of compliance. Importantly the Act prohibits transfer of personal data outside of Kenya without proof of adequate safeguards and consent of the data subject. For challenge participants requiring international data transfers time for approval as well as appropriate stakeholder engagement must be planned for in the challenge prize design.

Other key legislation regarding digital healthcare and data access include the *HIV and AIDS Prevention and Control Act (2006)* which guarantees individuals right to privacy, and details data confidentiality and any exceptions for disclosure of information. The *Health Act (2017)* which harmonizes fragmented legislation governing the health system structure, as well as regulations for the collection, processing, storage, use and disclosure of personal health information. The *Health Information Policy 2014-2030* guides the collection and processing of patient data, requiring de-identification of patient data before processing, and makes clear that health data should not be stored outside Kenyan territory.

Health data protection and privacy regulations are now well established in Kenya, and therefore so are the governance demands for challenge participants. Organisational capacity to meet these provisions in Kenya should be considered given the very recent introduction of the Data

\(^{143}\) Kenya Vision, 2030, [http://vision2030.go.ke](http://vision2030.go.ke/)
Protection Act with parts, such as the process for approval of data transfer safeguards, still in implementation. Low data management capacity and skills at primary care facilities, as well as fragmentation in the health system are also cited as a constraint on data use.\textsuperscript{144} Professional legal support, and ensuring sufficient time to address legal and capacity issues, should be budgeted for by challenge participants to access and use personal health data.

In terms of municipal services, the National Solid Waste Management Strategy (2014)\textsuperscript{145} by the National Environment Management Authority (NEMA) is the most recent government action establishing a common platform for action between stakeholders to systematically improve waste management. It introduces a new approach for improved waste management in Kenya to create wealth and employment and reduce pollution of the environment. However, the county government lacks operational capacity in both staff and equipment to undertake SWM in the city. The regulator NEMA is equally challenged and cannot effectively supervise SWM in the city.

Although digital literacy and data skills are a focus of government initiatives, health information system investment still remains low. Sustainability funding and end user engagement will support uptake and implementation of the tools designed in the challenge and may present opportunities for further funding within the Data for Science and Health programme by building scalable models across the country, or models that could be adapted for use in other areas of public health.

Full legal review for Kenya can be found in Appendix 17.

Uganda

In Uganda, the data protection law has already changed twice and does not seem to be widely adopted\textsuperscript{146}. However, the Ugandan government has identified ICT as central to achieving its Vision 2040 development strategy and has embraced e-governance as part of the implementation of its National ICT Policy (2014). In health, the Ministry of Health hosts a Central Data Bank sharing integrated health and health related information resources through a single knowledge management platform. Open data has also been promoted with Uganda undergoing a World Bank Open Data Readiness Assessment in 2015 and drafting an open data policy in 2017. An open data law though, has yet to be passed and while the Uganda National eHealth Policy (2016) covers certain aspects of open data, national policy fragmentation is often evident. The health data challenge prizes contribute to the government’s overall goals of innovative health data use though participants should ensure that stakeholders are aligned and engaged.

In 2019, Uganda passed the Data Protection and Privacy Act (2019) to implement the constitutional privacy provisions established by Article 27 of the Constitution of the Republic of Uganda (1995). The Act regulates personal data collection, processing, use and disclosure, and applies to any person, entity or public body within or outside Uganda, who collects, processes, holds or uses personal data. The provisions in the Act are in line with those applying to


\textsuperscript{145} National Environment Management Authority (2014), National Solid Waste Management Strategy, Nairobi.

\textsuperscript{146} Mark Okello’s interview
international standards (particularly GDPR) which should ensure ease of compliance. The processing or storage of personal data outside Uganda is permitted if adequate measures are in place in the country in which the data is processed or stored, at least equivalent to protections under the Act, or with the data subject consent. For challenge participants requiring international data transfers, time for approval as well as appropriate stakeholder engagement must be planned for in the challenge prize design.

The rights of individuals concerning use of their health data are set out in the Data Protection Act, the Patient’s Charter (2009) and the Medical and Dental Practitioners Act (1998). These include health data privacy and confidentiality, the right to access personal medical information, withdraw consent for personal data processing, and to have the data rectified, blocked and/or erased. The right to have decisions made by automated processing to be reviewed by a human is also established. The Health Commission Act (2001), defines the Uganda health system structure and includes provision that all health facilities have a designated person or committee responsible for observance of the individual’s rights. Further important guidelines for challenge participants to be aware of are the Uganda Medical and Dental Practitioner Code of Professional Ethics (2013) and the Uganda National Council for Science and Technology (UNCST) National guidelines for research involving humans as research participants, (2007).

Health data protection and privacy regulations are now well established in Uganda, and therefore so are the governance demands for challenge participants. Organisational capacity to meet these provisions in Uganda should be considered given the very recent introduction of the Data Protection Act. Low data management capacity and skills among healthcare workers, as well as inadequate or fragmented infrastructure in the health system, are also cited as constraints on data use. Professional legal support, and ensuring sufficient time to address legal and capacity issues, should be budgeted for by challenge participants for access and use of personal health data.

E-Health projects in Uganda have historically often remained as pilots and have stalled or stagnated following initial funding. Sustainability funding and end user engagement will support uptake, ownership and implementation of the tools designed in the challenge and may present opportunities for further funding within the Data for Science and Health programme by building scalable models across the country, or models that could be adapted for use in other areas of public health.

Full legal review for Uganda can be found in Appendix 18.

Malawi

Malawi does not have a comprehensive data protection and privacy legislation. Nonetheless, there are a number of legislations, regulations and policies that contain bits of provisions on data protection and privacy that impact on access, use or sharing of data. The Electronic Transactions and Cyber Security Act of 2016 is the closest legislation Malawi has on personal data privacy and protection and is therefore a key framework to take into consideration when designing the data access model for the challenge. It regulates how personal data should be

147 Kiberu et al (2017) Barriers and opportunities to implementation of sustainable e-Health programmes in Uganda: A literature review https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5458569/
processed to ensure the protection of data subjects, sets out principles which data collectors
must apply when processing personal data and lays out the rights of data subjects. This is
supported by The Constitution of the Republic of Malawi provides that every person shall have
the right to personal privacy, which shall include the right not to be subject to interference with
private communications including mail, and all forms of telecommunications. The Access to
Information Act 2016 provides for access to information held by the State and its organs as a
right for every citizen. However, it prohibits information holders from disclosing information
whose disclosure would result in the unreasonable disclosure of personal information about a
third party. The Medical Practitioners and Dentists Code of Ethics and Professional Conduct
also has relevant stipulations about patient consent regarding data sharing.

The National Statistics Act empowers the National Statistics Organization (NSO) to collect all
types of information, including personal information nationwide on behalf of the government.
They should therefore be included as a key stakeholder in the challenge as both a data provider
and potential beneficiary of any data infrastructure established.

In terms of capacity, one of the priority areas of the National ICT Policy of 2013 is human
capacity development through utilization of ICT to increase access to healthcare, education and
training facilities. This will also be a key policy to draw on in the design of the challenge in order
to ensure the activities are in alignment with policy objectives and gain governmental support.

In terms of IP, there is no legal framework that defines or provides for the general right of
ownership over data in Malawi. However, a general right of ownership over data can be found
in the National Healthy Information System Policy 2014 which provides that the ownership of
health data rests with the Ministry of Health.

Overall there are stakeholders, legislations and policies to draw on in the design and
implementation of this challenge, but they are not comprehensive and that entails a certain level
of risk. This risk can be mitigated through consistent legal support and advice, and a focus on
building strong stakeholder relations with key bodies such as the Ministry of Health. A full legal
review for Malawi can be found in Appendix 19.

Data Inventory

Below is a list of the data sources that would be relevant for this challenge, regardless of where
you ran it.

<table>
<thead>
<tr>
<th>Data type</th>
<th>Organisation/ Source</th>
<th>Availability</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surveillance</td>
<td>Pfizer Atlas (Antimicrobial Testing Leadership and Surveillance)</td>
<td>Available</td>
<td>14 years of cumulative data on changing bacterial resistance patterns across more than 60 countries.</td>
</tr>
<tr>
<td>Surveillance</td>
<td>WHO - GLASS</td>
<td>Available</td>
<td>National AMR surveillance systems, monitoring AMR trends and producing reliable and comparable data.</td>
</tr>
<tr>
<td>Surveillance</td>
<td>ODI / Wellcome - AMR register</td>
<td>Available</td>
<td>Industry sponsored human antimicrobial surveillance programmes.</td>
</tr>
<tr>
<td>Source</td>
<td>Description</td>
<td>Availability</td>
<td>Notes</td>
</tr>
<tr>
<td>---------------------------------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>--------------------</td>
<td>-------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Various sources</td>
<td>Antibiotic Resistance: Interdisciplinary Action (AR-IA)</td>
<td>To be sourced/gathered</td>
<td>Application developed through Wellcome’s Data reuse prize, developing an index of antibiotic resistance for common infection syndromes. Data was combined: data on antibiotic prescribing guidelines, aetiology of infections, access to and cost of antibiotics, antibiotic susceptibility from global antimicrobial resistance (AMR) surveillance datasets.</td>
</tr>
<tr>
<td>Burden of disease data</td>
<td>IHME - Global Burden of Disease (GBD)</td>
<td>Available</td>
<td>Global, regional, national, and in some cases subnational estimates of the burden of diseases, injuries, and risk factors. Visualisations available too.</td>
</tr>
</tbody>
</table>
| Prescribing data                      | IQVIA                                                                        | Only available through a corporate partnership | Framework of antibiotic prescribing reporting and analysis. Covering Ghana, Kenya, Malawi and Uganda:  
  - Isolate / subject number  
  - Isolate eligibility  
  - Isolate collection day (Note: this will not be a date, but will be designated as day 0 to support elements below)  
  - Isolate identification  
  - Isolate susceptibility testing results (site laboratory DST – drug susceptibility testing)  
  Patient  
  - Demography - age, sex  
  - Presence of comorbidities  
  - Presence/absence of clinical signs and symptoms of infection  
  - Was one or more of the following signs and symptoms of systemic inflammation present at the time of specimen collection?  
  - Fever ≥ 37.5°C  
  - White blood cell (WBC) count of >10,000 cells/mm3  
  - Local/regional lymphadenopathy  
  - Elevated C-reactive protein (CRP) (nonspecific)  
  - Site of Infection  
  - Infection treatment  
  - Hospitalisation status  
  - Mortality within the protocol window. |
<p>| Epidemiological and susceptibility data| Wellcome DRI team initiative                                                 | To be collected    |                                                                                           |
| Maps                                  | Global Research on Antimicrobial Resistance (GRAM) project                   | To be collected    | Maps demonstrating the AMR burden levels globally.                                        |</p>
<table>
<thead>
<tr>
<th>Behaviour data</th>
<th>Data source</th>
<th>Data collection status</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Behaviour data</td>
<td><strong>Drivers of Resistance in Uganda and Malawi (DRUM) project</strong></td>
<td>To be collected</td>
<td>Behaviours related to antibiotic use in Uganda and Malawi.</td>
</tr>
<tr>
<td>Behaviour data</td>
<td>Wellcome public engagement team - Citizen generated data</td>
<td>To be collected</td>
<td>Data on antibiotic acquisition and consumption, especially related to children (through a mix of qualitative and quantitative methods).</td>
</tr>
</tbody>
</table>
Urban health data challenge: Detailed Scoping

Disclaimer

The content of this challenge topic aligns closely with the strategic objective and direction of Wellcome’s Our Planet Our Health priority area. However, due to the likely impact of COVID-19 and the strategic restructuring in Wellcome in relation to environmental health, it is advised that the detail of this proposal is revisited in late 2020. In addition, Pathways to Equitable Cities Research partnership, funded by OPOH, will be publishing further work on phenotyping complex urban environments in late 2020 which can be built on in this challenge.

Summary recommendation

Key challenge question

How can we better understand and resolve public health risks in rapidly growing urban environments?

Why?

Urbanization is one of the leading global trends of the 21st century that has a significant impact on health. Over 55% of the world’s population live in urban areas, a proportion that is expected to increase to 68% by 2050\textsuperscript{148}. By 2030, projections indicate that two billion of the global urban population will live in slums, mostly in Africa and Asia\textsuperscript{149}. However, the health of people who live in slums is a topic that has received little attention\textsuperscript{150}.

A person’s risk of disease is affected by both personal factors, such as diet and their genetics as well as environmental factors such as sanitation and pollution, aka “neighbourhood effects”. Health authorities generally track information about individuals to inform decisions about public health and health services. But because of the concentration of people in slum areas, environmental factors affect a far greater number of people at any one time. Information about neighbourhood effects is therefore incredibly valuable for identifying highly impactful public interventions, but there is very little data about these communities.

For example, in many LMICs, open dumpsites are a common disposal method and in general, less than half of all solid waste in low-income countries is collected resulting in a significant

\textsuperscript{148} 68\% of the world population projected to live in urban areas by 2050, United Nations Department of Economic and Social Affairs, 2018

\textsuperscript{149} Ezeh, A., Oyebode, O., Satterthwaite, D., Chen, Y., Ndugwa, R., Sartori, J. The history, geography, and sociology of slums and the health problems of people who live in slums, 2016.

\textsuperscript{150} Slum health series, The Lancet, 2016
amount of illegal and unsafe disposal\textsuperscript{151,152,153,154,155}. Poorly managed solid waste has health, environmental, and economic effects that multiply as waste accumulates. Uncollected solid waste increases exposure of all individuals in communities to vector-borne and zoonotic infectious diseases carried by birds, insects, and rodents. Over time, uncollected waste accumulates to block waterways, resulting in flooding, contaminated surface and groundwater, and emissions of greenhouse gases like methane. Abuja is one of the fastest growing cities in the sub-Saharan Africa but the city is said to lack the modern management techniques to meet the requirement of a rapidly expanding city\textsuperscript{156}, often resulting to major floods that lead to significant deaths due to a “lack of proper town planning, blocked waterways and poor drainage systems\textsuperscript{157}.” Altogether, these neighborhood-level exposures lead to increased incidence of respiratory illness and diarrhea, and poorer mental health among individuals.

Municipal solid waste management is the largest budget item for city governments in most low-income and many middle-income countries and one of the largest employers\textsuperscript{158}. In LMICs, the amount of waste produced per person is expected to double in the next 20 years, and costs to manage solid waste will increase four to five fold\textsuperscript{159}. It is therefore in the interests of local government and policy makers in these regions to manage waste disposal better, but they have limited data about solid waste on which to base policies and allocate limited resources.

What?

We propose a data science challenge prize in areas experiencing rapid urbanisation such as Nigeria and Kenya that which closely engaged with city authorities to understand their specific municipal challenges, and the communities themselves to understand the effect of these challenges, it could motivate and support teams of data scientists working in the field of remote sensing to collaborate directly with public health experts. Through the combination of ‘remote sensing’ data such as satellite imagery, with administrative, municipal, health and community collected data, the challenge teams could develop models, tools and applications that support cities to make better decisions regarding public services that impact public health in urban areas. Due to the nature of ‘neighbourhood effects’, enabling cities to identify the specific interventions in slum areas that will have the biggest impact will support the health of a huge population, and help equip the authorities for dealing with their rapidly changing urban


\textsuperscript{153}Complex Urban Systems for Sustainability and Health research partnership


\textsuperscript{157}Nigeria floods kill more than 100, BBC, 2018, https://www.bbc.co.uk/news/world-africa-45546695


This research has identified that Nigeria and Kenya would be two suitable locations with a range of available data, networks of data scientists with sufficient capacity and initial interest from governmental authorities.

**Where?**

This document includes information on two countries: Nigeria and Kenya. These countries have been prioritised due to Wellcome’s strategic connections to Kenya, and the potential impact of the work on Nigeria.

ODI recommends that a challenge prize of this topic should be run in one country at a time, focusing on participants from across sub-Saharan Africa, though wider international applications could be considered. There are opportunities to scale the activities from one country to another, adjusting for specific needs and stakeholders.

**Who?**

<table>
<thead>
<tr>
<th>Creators (Participants)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>• Multidisciplinary teams with expertise in data science, environmental science, public health and software development could identify new insights and develop new solutions (analysis, products, tools) based on data combined for the first time.</td>
<td></td>
</tr>
<tr>
<td>• This expertise will be supported by people with skills in developing tools and services.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Decision makers and potential tools created to aid them</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>• Tools/ applications that can inform health and environmental policy-makers around policy priorities, for example using deprivation mapping across a number of cities to forecast state and national public health policy requirements.</td>
<td></td>
</tr>
<tr>
<td>• Tools/ applications that can be used by local authorities and private sector contractors to deliver better services related to urban health, for example enabling more effective and responsive solid waste management, sanitation or river pollution monitoring.</td>
<td></td>
</tr>
<tr>
<td>• Tools/ applications that support urban planning developments with neighbourhood level insights, for example enabling more targeted responses for rapidly urbanising areas.</td>
<td></td>
</tr>
<tr>
<td>• Tools/ applications which allow communities to advocate and respond to public health risks affecting their communities, for example the build up of solid waste in a way which is a flood risk.</td>
<td></td>
</tr>
<tr>
<td>• New tools for research that can help those without a data science background further explore the broader value of using remote sensing in combination with epidemiological data to understand public health risks and identify preventive mechanisms</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Beneficiaries</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>• Communities living in slum areas affected by public health risks</td>
<td></td>
</tr>
<tr>
<td>• City authorities who have responsibility for the services relating to health and the environment</td>
<td></td>
</tr>
</tbody>
</table>
Background Information

Urbanization is one of the leading global trends of the 21st century that has a significant impact on health. Over 55% of the world’s population live in urban areas, a proportion that is expected to increase to 68% by 2050\textsuperscript{160}.

Yet most of the 4.2 billion people living in cities still suffer from inadequate housing and transport, poor sanitation and waste management, and air quality failing WHO guidelines\textsuperscript{161}. The United Nations estimates that more than 90% of future urban population growth will be in low- and middle-income countries (LMICs).\textsuperscript{162} In sub-Saharan Africa, 62% of the urban population lives in slums\textsuperscript{163}. By 2030, projections indicate that two billion of the global urban population will live in slums, mostly in Africa and Asia\textsuperscript{164}. However, the health of people who live in slums is a topic that has received little attention\textsuperscript{165}.

An individual’s risk of disease is affected by both personal factors, such as diet and genetic constitution, environmental factors such as sanitation and pollution, aka “neighbourhood effects”. The health sector generally tracks individual-level indicators such as demographics, household income, nutrition or access to services. However, area-level or ecological indicators are often of greater use to decision-makers in setting priorities, allocating resources, and planning and evaluating development projects\textsuperscript{166,167}. These indicators combine individual level indicators with ‘neighbourhood’ level indicators which show things like levels of deprivation in relation to formal municipal boundaries and flood zones. See the Data Inventory section for an overview of indicators and relative data sources. There are increasing calls for interventions and programs addressing the relationship between individual-level health indicators and neighbourhood-level dimensions which shape health outcomes. In rapidly urbanising areas characterised by increasing populations living in informal settlements and slums, understanding the spatial context for health outcomes and interventions is critical.

This requires a way to identify slum-like neighborhoods and bring together data on their locations, spatial extents, demographics, and socioeconomic characteristics to allow for their adequate monitoring over time.

However, the people involved with urban health indicator development tend to have health and medical backgrounds and are unaware of, or are untrained in the use of, the types of data which measure neighborhood-level phenomena, such as satellite imagery. Further, the data scientists who work with such area-level datasets tend to be situated in the environmental

\textsuperscript{160} 68% of the world population projected to live in urban areas by 2050, United Nations Department of Economic and Social Affairs, 2018
\textsuperscript{161} WHO, Urban Health
\textsuperscript{162} WHO, Urban Health
\textsuperscript{163} Urbanisation, United National Population Fund
\textsuperscript{165} Slum health series, The Lancet, 2016
\textsuperscript{167} Ecological Models Revisited: Their Uses and Evolution in Health Promotion Over Two Decades
sciences or big industry with limited exposure to the ecological framework for health, and rarely package or distribute data with health decision-makers in mind\textsuperscript{168}.

For example, municipal solid waste management is the largest budget item for city governments in most low-income and many middle-income countries and one of the largest employers\textsuperscript{169}. In LMICs, the amount of waste produced per person is expected to double in the next 20 years, and costs to manage solid waste will increase four to five fold\textsuperscript{170}. In many LMICs, open dumpsites are a common disposal method and evidence suggests that in general, less than half of all solid waste in low-income countries is collected resulting in a significant amount of illegal and unsafe disposal\textsuperscript{171,172,173,174,175}.

Poorly managed solid waste has health, environmental, and economic effects that multiply as waste accumulates. Uncollected solid waste increases exposure of all individuals in communities to vector-borne and zoonotic infectious diseases carried by birds, insects, and rodents. Over time, uncollected waste accumulates to block waterways, resulting in flooding, contaminated surface and ground water, and emissions of greenhouse gases like methane. Altogether, these neighborhood-level exposures lead to increased incidence of respiratory illness and diarrhea, and poorer mental health among individuals.

It is therefore in the interests of local government and policy makers in these regions to manage waste disposal better, but they have limited data about solid waste on which to base policies and allocate limited resources. Example issues faced include problems such as:

- Inadequate information regarding the scale and impact of public health risks such as the build up of municipal solid waste
- Logistical issues in arranging contractors to effectively cover 100% of a city due to cross border dumping of refuse and public/private contractors operating simultaneously
- Inadequate facilities and limited resources to maintain and cover a whole area\textsuperscript{176}

Research indicates that mapping solid waste piles and estimating the volumes of rubbish they contain would be an enormous asset to those involved with solid waste management and


\textsuperscript{173} Complex Urban Systems for Sustainability and Health research partnership


\textsuperscript{176} Waste Management Strategies in Kaduna State MENR
planning in LMICs. Measurements of solid waste quantity and composition are generally taken at final dumping sites and via interviews with waste system managers, then supplemented with field visits to identify informal dumping sites and interviews with garbage pickers. However, the quality and completeness of these data vary substantially; they are altogether missing in many low-income countries.

Looking across epidemiological studies of the health impacts of exposure to solid waste, the literature points to a lack of ecological analysis and information on relevant confounders which limits public health researchers from being able to confidently evaluate the scale of impacts.

“Surveillance data are lacking due to the complexity involved in measuring exposure and outcomes but also the limited programmatic focus and funding to this area.”

Within the last decade, there has also been significant progress in the use of classical machine and deep learning approaches to map deprived areas. However, many of them work on relatively small areas and there are opportunities to expand the scale of their mapping to whole city areas. Data scientists using earth observation (spatial and satellite data) have manually identified and characterized dumping sites in small areas, and trained feature extraction models to identify dumping sites in large areas, though many of the latter studies focused on high-income countries. Research suggests that data scientists who wish to make substantial impact on health and wellbeing in LMICs should consider methods for mapping neighborhood-level health determinants such as solid waste pile location and coverage.

The combination of Earth Observation and GIS data can be used to predict human settlements, settlement type, and neighborhood outcomes such as total populations and population flows in areas as small as 100 x 100 m cells. Platforms such as OpenStreetMap make GIS data readily available, enriched by community mapping projects such as Map Kibera in Nairobi,

181 The Role of Earth Observation in an Integrated Deprived Area Mapping “System” for Low-to-Middle Income Countries
182 Impact of Inadequate Urban Planning on Municipal Solid Waste Management in the Niger Delta Region of Nigeria
alongside humanitarian GIS sources such as The Humanitarian Data Exchange and Map Action. Administrative boundary data is available through platforms such as GADM and DIVA.

This data can be combined with survey, census, administrative, and health system data to model data at the neighborhood-level with relevant accuracy, for example average household wealth by cell phone tower coverage area. WorldPop and ICF Interna-tional have already modeled dozens of household survey indicators in a gridded format, with estimated values for each small grid cell. Gridded datasets like these can be re-aggregated into meaningful geographic areas—for example, a city map of cultural neighborhood boundaries, city administrative wards, or health catchment areas—or viewed at the level of the city to get a sense of the distribution of health determinants such as sanitation, solid waste management or water pollution.

However, it is critical to take into account that slums are not homogenous and individual slums present very different scenarios in which a neighbourhood level intervention will play out. Therefore it is only through the combination of earth observation, remote sensing, and community based information and engagement that effective insights and interventions may be created.

These studies demonstrate that there is a significant amount of spatial data available to inform policies and services related to public health such as urban planning, transportation systems, supply of energy, water and sanitation, and waste management.

Kenya

The Lancet has reported that pollution in the air, water, and soil is responsible for 19.3% of all deaths in Kenya. In 2015, the report indicates that 58,000 Kenyans died from pollution-related disease, which is more deaths than from HIV, TB and malaria combined.

The APHRC’s 2000 Nairobi Cross-sectional Slums Survey showed that slum residents have the worst health and socioeconomic outcomes of any group in Kenya, including rural residents, with limited access to water and sanitation as well as education and employment. According to the World Bank’s 2016 review of urbanisation in Kenya, roughly 60 percent of Kenya’s urban households live in housing that would be defined as a slum under the Millennium Development Goals. Formal housing supply is not keeping pace with the growing urban population. Informal

---


193 The Role of Earth Observation in an Integrated Deprived Area Mapping “System” for Low-to-Middle Income Countries

194 Kenya & Pollution-Health and Economic Impacts, GLocal Alliance on health and Pollution. The Lancet
housing has become the only housing choice for most urban Kenyans. There are indications that households compromise on living conditions to remain within reasonable travel times of their jobs.

Population density in Nairobi varies greatly, peaking in the city's slums, which house roughly 2.5 million people in about 200 settlements\(^\text{195}\). Roughly 60 percent of Nairobi’s population occupies just 6 percent of the land, according to Kibera UK\(^\text{196}\). Along with the population growth and rapid urbanization, poor solid waste management is emerging as a key risk for health in Kenya\(^\text{197}\). The Dandora disposal site, Nairobi’s main dumpsite, is an open site and there are, in addition, more than 70 illegal dumping sites scattered throughout the city\(^\text{198}\). Though these challenges are most severe in informal settlements and communities living closer to dumpsites, they are not limited to these residential domains. Disposal is still primarily at the open dumpsite at Dandora, which becomes impassable during the rains, giving rise to illegal dumping in other parts of the city.

Tilahun Nigatu Haregu, Associate Research Scientist at the African Population and Health Research Center in Kenya has written that “health outcomes need to become a more central principle of solid waste management policy frameworks. As poor SWM is a critical risk factor for health, policies need to address health and the environment (or environmental health) in a more balanced manner.”

Nigeria

Nigeria’s population, according to the US Census Bureau is growing at a rate of 3.2% a year and is estimated will be at more than 400 million people in 2050 which will have enormous environmental health effects\(^\text{199},\text{200}\). More than half of Nigeria’s residents live in slum settlements, according to the 2015/2016 Slum Almanac.

Abuja is one of the fastest growing cities in sub-Saharan Africa but the city is said to lack the modern management techniques to meet the requirement of a rapidly expanding city\(^\text{201}\), often resulting to major floods that lead to significant deaths due to a “lack of proper town planning, blocked waterways and poor drainage systems\(^\text{202}\).”

Poverty and slum conditions pose a significant threat to urban health in the country. The current pattern of health investment seems to target the more affluent sections of the population who are the main users of the health services. Forced eviction and repression for urban poor are still common. Research indicates the need for targeted urban health interventions which meet the

\(^{196}\) Kibera Facts & Information, Kibera.org. Accessed 2.05.2020
\(^{200}\) The 100 million city: is 21st century urbanisation out of control?, The Guardian, 2018
\(^{202}\) Nigeria floods kill more than 100, BBC, 2018, [https://www.bbc.co.uk/news/world-africa-45546695](https://www.bbc.co.uk/news/world-africa-45546695)
needs of local authorities to respond to the pressures of rapid population growth, whilst also addressing some of the urban health risks\textsuperscript{203}.

Nigeria is also home to one of the world’s biggest floating slums, Makoko in Lagos. In response to sustained eviction threats, the community has recently undertaken a sustained mapping exercise using UAVs and Humanitarian OpenStreetMap\textsuperscript{204}.

### Theory of Change

A full theory of change is available in Appendix 1. The underlying hypothesis for the challenge is:

- **If we:** bring together teams of people with data science, remote sensing and urban health expertise with community organisers and city officials to work with a combination of geospatial, earth observation and big data in tandem with epidemiological data
- **Then:** the teams could use insights on neighbourhood level public health risk factors (eg: solid waste and pollution) in urban areas to create tools and products
- **Which:** could support a range of beneficiaries including community groups, city planners, public service delivery, national policy, academia and the development community to make better decisions and prioritisations for public services and health interventions in urban areas.
- **Ultimately leading to:** increased public health in that area and a peer network of experts who can continue to develop the data ecosystem

### Likely types of benefit

The challenge prize aims at developing insights and solutions for policy makers and researchers by bringing together environmental and public health expertise. With better identification of public health risks in urban areas such as solid waste build up, sources of river pollution, sources of hazardous medical waste which is not properly disposed of, the services and interventions needed to address these can be developed. In addition, making deprivation and slum mapping more efficient and accurate will aid preventative interventions can also be targeted due to an increased understanding of the likely risks in different urban areas, be they informal settlements, slums or more affluent areas. The solutions to the challenge prize might include models that identify solid waste from remote sensing data, or insights which demonstrate the health impact of different environmental factors through the combination of spatial and epidemiological data. Making the solutions scalable to multiple locations would also increase the types of benefit that may be created.


\textsuperscript{204} How Makoko, Nigeria’s floating slum went digital with new mapping project, CNN, 2020
Success indicators

**Impact indicator:** Improved public health in that urban area due to better town planning and public services.

A set of specific health indicators will be selected in consultation with local health facilities as a proxy measurement for improved public health. Two options will then be considered for collecting corresponding data to track progress at the local population level:

1. Local health facility data in line with existing practices to monitor the implementation of health initiatives at sub-national level.\(^{205}\)
2. Data available through the District Health Information Software 2 (DHIS2) for Kenya\(^{206}\) and Nigeria\(^ {207}\).

These options will be considered based on the availability, reliability and relevance of datasets for the urban areas selected for the challenge.

**Mid-term outcome indicator:** Use of tools and products created during the challenge by town planners and public service providers to identify policy and service interventions to prevent or reduce the public health risks posed by environmental factors.

Additionally to these success indicators, this challenge prize will meet wellcome’s intended aims for the challenge prizes in the following ways:

- **Increasing public trust:** community members and residents will be part of the co-design process of the challenge in that area, ensuring that their needs are taken into account alongside the needs of the city officials and research community. Representatives will also be part of the judging panels and maps and data assets generated through the prize will be shared with these communities in digital and print forms so they can be repurposed for other uses.
- **Increasing capacity:** Research suggests\(^ {208}\) there is considerable value in bringing together environmental scientists with data science expertise with public health officials. As training and mentorship will happen throughout the data challenge programme, making sure participants have the capacity to take part and build long term data science capacity that will benefit the city and country in question.

Potential time to impact

Depending on the engagement with key beneficiaries in local authorities, there is potential for relatively rapid impact (1-2 years after launch of tools) if the models are used to clear public health risks in the short term. However, the greater impact will be seen in the long run through a better understanding of the ecological factors of public health in urban environments and longer term town planning initiatives.


\(^{206}\) [https://hiskenya.org/dhis-web-commons/security/login.action](https://hiskenya.org/dhis-web-commons/security/login.action)


Strategic alignment with Wellcome

Wellcome have an ongoing commitment to environmental and planetary health. This proposal aligns directly with the Our Planet Our Health priority area focus on urban health and aligns with two key Wellcome funded programmes: Complex Urban Systems for Sustainability and Health research partnership in Kenya and Pathways to Equitable Healthy Cities in multiple locations. Pathways to Equitable Cities Research partnership will be publishing further work on phenotyping complex urban environments in late 2020 which can be built on in this challenge.

Nigeria is a region which Wellcome has not historically done a huge amount of work in, but given the extent of networks and resources identified through this research this could be an opportunity to engage with the region.

Over the next 12 months, Wellcome will be defining the Environmental Health strategic pillar and therefore it is recommended that this topic is reviewed in late 2020 to ensure full alignment.

Data landscape summary

In general, many LMICs do not have a significant amount of data infrastructure to monitor the relationship between the built environment and public health outcomes. There are many global health indicators, but they are rarely used in LMICs. Data which does exist is usually out of date which is why there is a need to use alternative data sources to track key indicators. For services such as municipal solid waste collection, human and resource capacity to carry out these studies which involves the collection of informative data on waste composition and quantity that is hauled to treatment sites or recycling centers or disposal sites is lacking. Town planners, environmental regulators and local authorities lack the information they need on urban health risks and their relation to key municipal services in order to plan city services and interventions more strategically.

However, there are 500 health indicators of the physical environment which can be used to inform public health decision-making in LMICs. In each of these efforts, indicator identification was necessarily constrained by available datasets—those typically considered relevant include household surveys such as:

- Demographic and Health Surveys (DHS)
- censuses
- administrative records
- health system data
- national and sub-national policy documents
- household surveys

All of the above are available in each of the countries evaluated in this paper (Nigeria and Kenya) though some of the data is out of date.

---

200 See Data Inventory for further detail.
Datasets suitable for the measurement of small areas are needed to calculate neighborhood-level determinants, including data collected by

- Earth Observation (EO)
- Geographic Information Systems (GiS)
- big data (e.g., mobile phone records, traffic sensors, pollution sensors)
- climate data
- field observation of areas (not households)
- traffic sensors (to measure air pollution)

The spatial resolution provided by satellite imagery is sometimes not sufficient for e.g., the detection of individual houses, infrastructure, and details of environmental conditions. Unmanned Aerial Vehicles (UAVs), also known as drones or Remotely Piloted Aircraft Systems (RPAS), are a potential solution for this issue.

In addition it is vital that insights from remote sensing sources can be ground truthed and cross referenced with community collected data. This also has the benefit of ensuring engagement with the residents of the areas in question. Slum Dwellers International do annual community mapping exercises while Data Science Nigeria has extensive experience in ground truthing. This could be done through drawing on the Surveys for Urban Equity project manual for area observation surveys\(^\text{210}\), based on the previous UN-Habitat manual. This manual has been piloted but not scaled up so the opportunity to create larger observational data sets would be valuable.

**Challenge structure**

The challenge period will cover a period of 18 months, with a 9 month set up phase preceding the prize, and 9 month sustainability prize following the judging.

As previously mentioned, this challenge will need to be run in a single country, with a specific set of end user beneficiaries and community representatives which need to be engaged with throughout the design and delivery of the challenge prize. Participants should be recruited from across the country surrounding the region, and supported to travel to key events. This is intended to enable a peer network of local experts across the region the challenge is being run in.

Seed funding would be needed in order to allow for people to form multi-disciplinary teams and take part in the competition. Details of this are available in the budget and in the descriptions below.

Training and mentoring would also be required, to align capacity and to ensure benefits would be lasting for the participants and the country as a whole. Specific training requirements will be established early in the challenge, depending on what participant needs are. Training might

---

cover relevant areas from the data skills framework, such as ethical use of data or machine learning approaches.

Remit and scope of the challenge

Key question
How can we better understand and resolve public health risks in rapidly growing urban environments?

Suitable Proposals
Proposals should be submitted to make progress in one or more of the following areas:

- Tools/applications that can inform health and environmental policy-makers around policy priorities, for example using deprivation mapping across a number of cities to forecast state and national policy requirements for public health.
- Tools/applications that can be used by local authorities and private sector contractors to deliver better services related to urban health, for example enabling more effective and responsive solid waste management, sanitation or river pollution monitoring and forecasting the potential impact on health.
- Tools/applications that support urban planning developments with neighbourhood level insights, for example enabling more targeted responses for rapidly urbanising areas with significant levels of informal settlements.
- Tools/applications which allow communities to advocate and respond to public health risks affecting their communities, for example the build up of solid waste in a way which is a flood risk.
- New tools for pre-clinical research that can help researchers without a data science background further explore the broader value of using remote sensing in combination with epidemiological data to understand public health risks and identify preventive mechanisms

Any proposal submitted should consider the end goal of enabling health and environmental authorities and urban communities to identify and respond to environmental health risks.

Who can apply
We are looking for multidisciplinary teams who bring data science together with a range of alternative disciplines in order to deliver innovative solutions to this complex issue. Example disciplines include:

- environmental science
- public health
- remote sensing
- computer science and machine learning
- digital development
- service design

The prize will be open to all international participants, but there will be significant engagement and recruiting with local networks of experts and all events will be run in the location specific to the prize.
The application process is open to those who have predefined ideas and projects, and those who may have relevant skills but no project proposal.

Participant motivations

The motivation for participants to enter will be a combination of:

- opportunities for networking / building relationships
  - access to key government and healthcare sector stakeholders
  - access to other experts within and across disciplines
  - access to commercial partners and clients
- personal and professional development
  - advancing the field of snakebite research
  - advancing technological development
  - advancing individual capabilities
- personal and professional rewards
  - financial reward of winning the prize
  - reputational reward of winning the prize

The motivation for academic participants tends to be more driven by access to key government and healthcare stakeholders, improvement in and contributions to their field of study, and reputational awards.

The motivation for commercial participants tends to be more driven by access to new markets and clients, positive press around service provision and financial rewards.

The motivation for governmental stakeholders will be to outsource the development of innovations that can help them deliver more efficient services, and contribute towards SDG goals.

Locations and participant recruitment

ODI's recommendation is that while both countries present many opportunities, the challenge should be delivered in a single country, not both simultaneously. Therefore Wellcome should use the information in this report along with further consultation with stakeholders to prioritise one country over another (Kenya or Nigeria). It would be possible to run the same challenge sequentially first in one country, and then the other. This has not been accounted for in the recommended sequencing or budget, however it is feasible.

Due to the nature of the challenge focusing on urban areas, during the set up phase the delivery team is advised to work with government stakeholders to identify one or several key cities in the chosen country to work with. The more cities which are included in scope, the more complex the stakeholder liaison and data infrastructure will be. However, the more cities, the greater the likely variation of interesting innovations due to the richer data, greater use cases, and opportunities for comparison. The engagement from municipal and governmental stakeholders is crucial to ensure the correct focus, the right data is made available and the uptake of any solutions developed is guaranteed.

In terms of recruitment for participants, there are extensive domestic data science networks across these regions, such as Data Science Africa and Machine Intelligence Africa in addition to
country specific networks such as [Data Science Nigeria](#) and the [Geoinformation society of Nigeria](#). The recommendation therefore is to focus on these networks for recruitment. This will have the consequence of building on the existing links across the networks to enable a pan-African support network.

**Challenge stages, funding & support**

The estimated time allocations for the mental health data challenge would include:

- Up to **9 months** for a set up phase
- Up to **18 months** for the challenge to run, from recruitment of participants through to the awarding of the challenge prizes
- Up to **an additional 9 months** for a sustainability period

For a detailed view of estimated timelines, please refer to the urban health data challenge timeline in Appendix 10.

1. **Set up phase**

   The organisation that is selected to run this challenge programme would likely need six to nine months in order to set up the challenge programme. This might include, but is not limited to, the following activities:

   1. Convening beneficiaries and stakeholders across the communities in question, local, state and national government and private sector in order to verify their needs with regard to urban health and municipal management and confirm their engagement with the prize.
   2. Convening community representatives to understand how municipal needs relate to community needs.
   3. Establish agreements with local delivery partners, and establish the detailed delivery plan.
   4. Establish agreements with data providers. Depending on the available data in the chosen location, some ground truthing/community data may need to be collected prior to launch.
   5. All data will need to be vetted and cleaned thoroughly prior to launch in order to verify quality and limitations. This will take dedicated resources and be the most
time consuming part. The aim is to ensure that participants can use the data immediately at the point they enter the Insights Phase.

2. Participant recruitment/ Expression of interest
A period of up to three months will be required for recruitment of participants. This should overlap with the set up phase. A website will need to be established with all the required information on topic and processes. This will then be publicised through some advertising and media channels. In this case, it is likely that the bulk of the participants will be recruited through established networks such as the African Association of Remote Sensing if the Environment or the Geoinformation Society of Nigeria.

Participants will be invited to complete an expression of interest (EOI). EOIs can be open to those who have predefined ideas and projects, and those who may have relevant skills but no project proposal. Participants will be able to communicate with each other during this period, and will be encouraged to make connections and discuss ideas to help with team formation.

3. Launch lab
The challenge prize will begin with a ‘lab’ or multi-day event (possibly virtual given the Covid-19 crisis). All those who have completed an EOI will be invited, along with beneficiaries, data providers and delivery partners. Participants will have their expenses paid for travel and accommodation.

The first part of this event will focus on explaining why this challenge topic is so important, telling the stories of why and how this problem needs to be addressed, and the processes involved in participating in the challenge prize. This can be recorded to be sent to people who cannot attend. Press will be invited to cover the ‘launch’ and meet some of the participants and key stakeholders.

Following this introductory session, participants will have the opportunity to share their project ideas, recruit new team members, and engage with data providers and stakeholders.

The aim is to support the development of multidisciplinary teams, bring them into contact with the data providers and beneficiaries, and kick start the development of their proposals and ideas through two days of workshops and activities. There will be a clear process for how to team up and formally enter as a team.

4. Team applications
The teams will be given 4-8 weeks following the launch lab to refine their proposal ideas, confirm their teams and submit their area of interest/ initial proposal in a formal team application. This will then be assessed based on criteria including:
- Multidisciplinary team with relevant expertise
- Area of interest in line with challenge scope
- Scale of ambition

During this period, participants will have access to a challenge team application template to complete, have access to supporting information on the topic, and will be able to troubleshoot with the organisation running the challenge for process queries.
Teams may submit their proposal at any time during this period and begin the next phase, though there will not be access to funding before 4 weeks. The application window will remain open until 8 weeks after the launch lab.

5. **Model development phase**

   This phase will be between 6-9 months and focus on training and the development of data insights. The teams will be encouraged to focus on engaging with the available data and modeling it to develop insights. During this time they will receive some seed funding (£25k for 10 groups), data access support to help them identify the data they need for their project and any governance requirements, and a rolling series of seminars on relevant machine learning techniques, public health and environmental management.

   In this phase we would expect the members of the teams with data science, public health and environmental science to be leading the work.

   Through the development of models and insights, the teams will be asked to put together plans for developing practical applications and tools for end-users drawing on the data science methods and insights they have developed. These tools and applications will need to address specific needs of beneficiary groups such as, but not limited to:

   a. Tools/ applications that can inform health and environmental policy-makers around **policy priorities**, for example using deprivation mapping across a number of cities to forecast state and national policy requirements.

   b. Tools/ applications that can be used by local authorities and private sector contractors to **deliver better services related to urban health**, for example enabling more effective and responsive solid waste management, sanitation or river pollution monitoring.

   c. Tools/ applications that support **urban planning developments with neighbourhood level insights**, for example enabling more targeted responses for rapidly urbanising areas with significant levels of informal settlements.

   d. Tools/ applications which allow **communities to advocate and respond** to public health risks affecting their communities, for example the build up of solid waste in a way which is a flood risk.

   e. New tools for pre-clinical research that can help **researchers without a data science** background further explore the broader value of using remote sensing in combination with epidemiological data to understand public health risks and identify preventive mechanisms.

The model development phase will culminate in a demo day in which all of the teams will demo the models and insights they have developed using the data, and their pitches for tool development. These will be judged by a panel of key stakeholders and experts in the field against a set publicly accessible scoring criteria. For example, applications might be assessed on some of the following areas with a focus on evaluating the innovation of the data modeling done so far, the feasibility of the proposal and the potential impact:
• Understanding of the problem
• The proposed approach
• Team resources
• Proposed outcomes
• Potential impacts
• Project management documentation
• Risk assessment
• Cost plan
• Ethical assessment

The best teams will be selected to go through to the development stage to build their proposed tools/applications. The number of successful applicants who move from the model development phase to the tool development phase would likely be pre-agreed at this stage, in order to control costs in the discovery phase. Alternatively applications could be judged on merit, but this would mean there is no limit to the number of participants in the first phase and would cost more to run.

6. Tool development phase
The tool development phase will also run for between 6-9 months. The projects that have progressed will receive further seed funding (£100k for up to 5 teams) in order to develop the tool or application outlined in the insights phase.

In this phase we would expect the teams to have a greater focus on digital development, working with and drawing on the expertise of the team’s scientific experts and stakeholders to inform their work. These teams will receive training and mentorship suitable to their specific topic to enable their ideas to develop and flourish. This will include support to plan implementation with stakeholders, ethical assessments and business model development.

Where possible, the challenge teams will get a chance to implement their full or partial solutions in an environment with real stakeholders. This will allow for feedback, adjustments and further development before the final judging. This will need to have been agreed in principle with key stakeholders during the implementation phase, and then coordinated by the delivery team following the insights period.

7. Judging
The development phases will culminate in a public demo day attended by as many media representatives and senior stakeholders attending as possible. Press and marketing will be a key part of this, with the prize event aiming to give the programme and the winners promotion and visibility. Each of the teams present their work. The event will have a strong focus on storytelling - what the nature and scale of the risks around urban public health are, and how we have used data science to address them.

A set of judges who are known experts in the field, such as Dana Thompson, Jose Sri, and representatives from the group of beneficiaries such as NEMA (Kenya) or NASRDA (Nigeria), will review the submissions against publicly available criteria. Criteria are likely to include:

• Innovation
  ○ Ethical data management
  ○ Utilises multiple data sources

• Impact
  ○ Scale of potential impact on urban health in the region
- Building trust with the public on the use of data for health innovation
- Capacity
  - Clear input from multiple disciplines
- Beneficiaries
  - Clear link to government/other appropriate end user
  - Scalable solution which has the potential to be used in other regions
- Sustainability
  - A clear, time bound implementation plan for the solutions to be adopted by end users
  - Whether the solution have the potential to continue beyond the prize

The most successful submission will be awarded a prize of £200k. Three runner up teams shall receive £100k. This money will be explicitly for the implementation and scaling of the tool they have developed. The rationale for this model is that it allows multiple tools and outputs to go on to be scaled and supported, delivering sustainable impact in the field. This has been benchmarked against other health technology innovation prizes (see Appendix 8).

8. **Sustainability/ Uptake period**

It is vital that there is a sustainability period following the award of the prize in order to support the winning project and the other projects to be adopted by the key stakeholders and used to address the key issues addressed regarding mental health. The winning teams can use the prize money to implement and scale up the innovations they have developed.

In addition, the partners involved in the delivery of the prize will need to remain available for the teams over a period of 9 month in order to:
- Assist the future development of winning solutions by making introductions to would-be users, investors, commissioners and others.
- Establish a schedule for checking progress with winners and interesting other participants.
- Seek feedback from competitors, judges and partners and evaluate the effectiveness of the programme.
- Return to the prize winners a year afterwards to develop a case study of where they have got to in their development and the impact they have created.

9. **Evaluation**

At the end of the judging and sustainability phase there shall be an independent evaluation made of the prize as a whole drawing on reporting material and interviews with participants and stakeholders. The interviews will also be used to develop case studies of each output and prize in order to share learnings and successes for others interested in the field of health data innovation and remote sensing for public health.
Challenge prize data ecosystem

An ecosystem map can be helpful in understanding how data flows between different stakeholders involved in snakebite management and how value is created in the ecosystem by allowing different stakeholders to make better decisions.

To produce the below diagram, We used the ODI's data ecosystem mapping tool. It is a visual illustration of how data is being accessed, used and shared.

**Urban Health Challenge Data Ecosystem Map.** A high res PDF version of the map is available in Appendix 14 and an interactive version is available [here](#).

### Risks

There are three major risks inherent with this challenge topic. The first is lack of engagement form and with local authorities. Without full engagement there is a high risk that any solutions produced will not be adopted and therefore not deliver the desired impact.

The second major risk is around lack of engagement with the communities themselves in the areas under consideration. It is vital that communities are not treated as homogenous and the specific cultural variables and sensitivities are taken into account. This will make the solutions more effective, and go towards minimising any potential undermining of the communities themselves. Ensuring that all data is balanced with ground truthed data, and the community representatives are involved in the design and delivery of the prize is crucial.
Lastly the use of remote sensing raises the risk of state surveillance. Many urban poor communities are at risk from violence and eviction so ensuring that any solution created does not put them at risk is of high importance. Engaging with the communities as well as legal and ethical experts is one way of minimising this risk.

A full risk register is available in Appendix 3.

### Capacity and stakeholder landscape

#### International Delivery partners

The following organisations could be relevant for engaging in the delivery of this data science challenge prize across Africa, not just in a single country. Country specific partners are listed in the regional sections.

<table>
<thead>
<tr>
<th>Organisation</th>
<th>Description</th>
<th>Role</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>African Population and Health Research Centre</td>
<td>Africa’s premier research institution and think tank, generating evidence to drive policy action to improve the health and wellbeing of African people.</td>
<td>Academic advisor</td>
<td>TBC</td>
</tr>
<tr>
<td>Slum Dwellers International</td>
<td>A network of community-based organizations of the urban poor in 33 countries in Africa, Asia, and Latin America including Nigeria and Kenya.</td>
<td>Data provider and delivery partner</td>
<td>The secretariat of all SDI has stated they are interested in partnering</td>
</tr>
<tr>
<td>Data Science Africa</td>
<td>Data science community operating across sub-Saharan Africa</td>
<td>Delivery partner</td>
<td>Interested in partnering</td>
</tr>
<tr>
<td>MapBox/ Open Street Map</td>
<td>An open source mapping platform for custom designed maps.</td>
<td>Delivery partner</td>
<td>Interested in partnering</td>
</tr>
<tr>
<td>Facebook Map With AI</td>
<td>In conjunction with OpenStreetMap, using artificial intelligence to predict features on high-resolution satellite imagery.</td>
<td>Technology partner</td>
<td>Interested in partnering</td>
</tr>
<tr>
<td>Grid3</td>
<td>GRID3 (Geo-Referenced Infrastructure and Demographic Data for Development) works with countries to generate,</td>
<td>Technology partner / data provider</td>
<td>Interested in partnering</td>
</tr>
</tbody>
</table>
validate and use geospatial data on population, settlements, infrastructure, and subnational boundaries.

<table>
<thead>
<tr>
<th>Organization</th>
<th>Description</th>
<th>Role</th>
<th>Interest</th>
</tr>
</thead>
<tbody>
<tr>
<td>University of Twente, Department of Urban and Regional Planning and Geo-Information Management</td>
<td>Leading academics in the use of remote sensing for urban health</td>
<td>Academic advisor</td>
<td>Interested in partnering</td>
</tr>
<tr>
<td>Pure Earth: Blacksmith Institute</td>
<td>Pure Earth is an international non-profit organization dedicated to solving pollution problems in low- and middle-income countries, where human health is at risk.</td>
<td>Delivery partner</td>
<td>TBC</td>
</tr>
<tr>
<td>Machine Intelligence Africa</td>
<td>Machine Intelligence Institute Africa</td>
<td>Delivery partner</td>
<td>TBC</td>
</tr>
<tr>
<td>African Association of Remote Sensing of the Environment</td>
<td>Earth Observation advocacy group</td>
<td>Delivery partner</td>
<td>TBC</td>
</tr>
</tbody>
</table>

Kenya

**Governmental stakeholders and data providers**

Kenya’s [National Environment Management Authority](#) (NEMA), the department which has oversight of solid waste, pollution and flood management is interested in partnering with Wellcome on a challenge prize of this topic. Furthermore they have agreed to act as a convener with other key governmental stakeholders in order to facilitate access to data from the Ministry of Health and National Bureau of Statistics. In addition NEMA have indicated that the [Nairobi Regeneration Group](#) would be an ideal group of end users with whom participants could collaborate.

**Data Science**

- Data Science programmes at University of Nairobi and [Moringa School](#)
- Kemri/ Wellcome has significant expertise in health modelling and machine learning, though it is usually focused on Malaria
- [Data Science Nairobi](#) is a non-profit professional group that brings together people interested in data science in Nairobi with over 1500 members.

**Complex Urban Systems for Sustainability and Health research partnership**

[CUSSH](#) is a Wellcome funded initiative related to the OPOH priority area. They work extensively in Kenya and support decision makers in linking health and environmental policy. For this initiative they could provide facilitation for the links with government stakeholders.
APHRC/ Nairobi Urban Health and Demographic Surveillance System (NUHDSS)
This system is a pioneer urban health data surveillance system established was established in two slum communities in Nairobi (Korogocho and Viwandani) by the APHRC. The platform routinely collects data on demographic events (births, deaths, and migrations), health outcomes (morbidity, cause of death through verbal autopsy, child vaccination, and nutrition) and socio-economic outcomes (marriage, education, livelihood, and housing characteristics). Additionally, the NUHDSS also serves as a platform for other specialized studies focusing on urbanization, population, education, family planning, reproductive and general health conditions of the urban poor. The data is openly available.

SLUMAP
This extensive project looking at slum mapping in Kenya through remote sensing could provide an excellent data and conceptual framework to be utilised for public health purposes. They will have delivered the research by 2021.

Ground truthing
Map Kibera is a very established community mapping initiative which works across Nairobi to support slum mapping based on Open Street Map. Slum Dwellers International has a Kenyan Federation which also does annual data collection on slum communities across the country. Both organisations have been engaged with and are interested in supporting this initiative and open to specific data collection in order to provide data on specific community issues.

Nigeria

Governmental stakeholders and data providers
The National Space Research and Development Agency (NASRDA) who hold all national geospatial data as well as the Geoinformation Society of Nigeria (GEOSON) the umbrella body for Geoinformation practitioners, academia and researchers in Nigeria are interested in partnering with Wellcome on an initiative combining remote sensing with public health initiatives.

They could provide a number of highly valuable datasets, including access to the NigeriaSat-2 and Ortho-rectified Spot-5 satellite data. In addition, NARSADA is in the process of taking over the ownership and maintenance of the Grid3 Nigeria georeferenced data infrastructure which includes over 980 publicly available data sets in 37 states relating to administrative boundaries, population, energy, settlements, health, facilities and others. Grid3 are also keen to partner with Wellcome on this initiative and would be able to provide local expertise in governmental liaison and geospatial data.

Dr. Matthew Adepoju who is president of the Geoinformation Society of Nigeria and Head of the Strategic Space Applications Department of NASRDA is keen to collaborate and has agreed to convene these wider stakeholders, in addition to engaging with the wider geoinformatics community. His team are also in the process of completing a needs assessment survey across all the ministries and waste management authorities across the country to identify common issues and can be shared with Wellcome on completion in order to support the implementation of this challenge. Other key stakeholders include:

- Lagos Waste Management Authority (LAWMA)
- Lagos State Environment Protection Agency
- State Ministries of Environment Lagos State Waste Disposal Board (LSWDB)
Data Science Networks

Data Science Nigeria is a leading data science community of practice who are keen to partner with Wellcome on this. They would be an ideal network through which to recruit participants, but potentially also to manage delivery of the challenge prize. In addition they have a strong background in ground truthing so if there are additional data collection requirements they would be prepared to undertake this.

Slum organizations (community engagement and ground truthing)
- Black Diamonds Support Foundation
- Shack/Slum Dwellers International (SDI)
- https://www.justempower.org/about-jei
- Nigeria Slum/Informal Settlement Federation
- Slum2School
- iBelieve Foundation

Local NGO on environment/ water/ sanitation (potential end users)
- Just Empower
- AnAwake Foundation
- The Nigerian Environmental Society
- The Centre for Climate Change and Environmental Studies
- Friends of The Environment
- Recycle Points
- JDL Charity
- Water Initiatives Nigeria (WIN)

Public Sector Waste Management
- The Association of Waste Managers of Nigeria (AWAMN)
- Waste Management Society of Nigeria (WAMASON)
- LBS sustainability centre
- ACT Foundation
- E-waste Producer Responsibility Organisation Nigeria, EPRON

Private Sector CSR
- Dangote Waste to Wealth Initiative
- Dreg Waters Petroleum & Logistics
- Fidelity Bank CSR for Environment

Wider Challenge Landscape

City Alliance’s Cities Without Slums Action Plan
Cities Alliance is a global partnership fighting urban poverty and supporting cities to deliver sustainable development. To manage its activities, the Cities Alliance operates a multi-donor fund with UNOPS as host and Trustee. In 1999, at the inaugural meeting of the Cities Alliance, Nelson Mandela launched Cities Alliance’s ‘Cities without Slums’ 20 year action plan with the goal of achieving a significant improvement in the lives of at least 100 million slum dwellers by 2020. They have specific programmes of work in Liberia, Tunisia and Uganda.

The Global Alliance on Health and Pollution (GAHP)
GAHP is a collaborative body made up of more than 60 members and dozens of observers that advocates for resources and solutions to pollution problems. GAHP was formed because international and national level actors/agencies recognize that a collaborative, multi-stakeholder, multi-sectoral approach is necessary and critical to deal with the global pollution crisis and resulting health and economic impacts. GAHP’s overall goal is to reduce death and illness caused by all forms of toxic pollution, including air, water, soil and chemical wastes especially in low and middle-income countries. This is done through advocacy and country specific actions plans.

**WHO Urban Health Initiative (UHI)**
The WHO Urban Health Initiative (UHI) aims to reduce deaths and diseases associated with air and climate pollutants, and enhance health co-benefits from policies to tackle urban air pollution and short-lived climate pollutants (SLCPs) – saving lives by linking health, environment and sustainable development. The Urban Health Initiative aims for cities to have the data, tools, capacity and processes to include health in the development equation.

**WHO Urban HEART Tool**
The Urban Health Equity Assessment and Response Tool (Urban HEART) is a decision-support tool to identify and reduce health inequities in cities. It has been implemented in over 40 countries and enables local communities, programme managers, and municipal and national authorities to:

- better understand the unequal health determinants, unequal health risks and unequal health outcomes faced by people belonging to different socioeconomic groups within a city (or across cities);
- use evidence when advocating and planning health equity interventions;
- participate in intersectoral collaborative action for health equity;
- apply a health equity lens in policy-making and resource allocation decisions.

**Centre for sustainable, healthy and learning cities and neighbourhoods**
The GCRF Centre for Sustainable, Healthy and Learning Cities and Neighbourhoods (SHLC) is an international consortium of nine research partners aiming to strengthen capacity to address urban, health and education challenges in neighbourhoods across fast-growing cities in Africa and Asia. Our partners include: University of Glasgow, Human Sciences Research Council, Ifakara Health Institute, Khulna University, Nankai University, National Institute of Urban Affairs, University of Rwanda, University of the Philippines Diliman, University of the Witwatersrand.

**The International Society for Urban Health**
The International Society for Urban Health (ISUH) is a global organization committed to making cities healthier by addressing the broad determinants of health and health inequities. The New York Academy of Medicine (the Academy) is the secretariat of ISUH. Their work covers the translation of science to policy, testing and scaling interventions, and support multilateral research programmes. The majority of the work focuses on adolescent health in Latin America.

**The Earth Institute**
Based at Columbia University, the largest research unit is the Lamont-Doherty Earth Observatory. The institute creates multidisciplinary collaborations that help us learn how to best address issues of global sustainability. They use that knowledge to develop policy and engineer practical solutions to climate change and environmental degradation, rapid urbanization,
infectious diseases, natural hazards, the need for clean and accessible energy, water and sanitation, and the sustainable use of resources.

**Pathways to Equitable Healthy Cities**

Funded by the Wellcome OPOH priority area, PECH is a global partnership that aims to improve population health, enhance health equity and ensure environmental sustainability in cities around the world through co-production of rigorous evidence with policy and civil society partners in multiple countries. They use diverse data sources, especially emerging open and big data, and novel methods for data integration and visualisation to characterise cities dynamic social, physical and natural environments, people’s experiences of these environments, and their health. Building on their work and connections would be of strategic benefit to Wellcome. They will be publishing further work on phenotyping complex urban environments in late 2020 which can be incorporated into this challenge.

**Legal, policy and regulatory environment**

**Kenya**

The Kenyan government has driven the growth of a significant data economy across both public and private sectors as part of its goal of hosting a globally competitive knowledge based economy by the year 2030. In 2011, Kenya launched the Kenya Open Data Initiative, the first such programme in Sub-Saharan Africa, making key government data, including in health, freely available through a single open data portal, [www.opendata.go.ke](http://www.opendata.go.ke). Further activities include the launch in 2016 of the Kenya Health Data Collaborative, an initiative with global health partners to improve the availability, quality and use of data for decision-making. In this context, the health data challenge prizes align with the government’s commitment to enabling the innovative use of health data.

In 2019, Kenya passed its Data Protection Act implementing Article 31 of the Constitution of Kenya (2010), the right to privacy. The Act harmonized the legal framework on data protection, becoming the primary and comprehensive statute. The Act defines key terms and roles, and makes clear requirements for express and informed consent for the collection and processing of personal data. The provisions in the Act are similar to those applying to international standards (particularly GDPR) which should ensure ease of compliance. Importantly the Act prohibits transfer of personal data outside of Kenya without proof of adequate safeguards and consent of the data subject. For challenge participants requiring international data transfers time for approval as well as appropriate stakeholder engagement must be planned for in the challenge prize design.

Other key legislation regarding digital healthcare and data access include the HIV and AIDS Prevention and Control Act (2006) which guarantees individuals right to privacy, and details data confidentiality and any exceptions for disclosure of information. The Health Act (2017) which harmonizes fragmented legislation governing the health system structure, as well as regulations for the collection, processing, storage, use and disclosure of personal health information. The Health Information Policy 2014-2030 guides the collection and processing of

---

211 Kenya Vision, 2030, [http://vision2030.go.ke](http://vision2030.go.ke/)
patient data, requiring de-identification of patient data before processing, and makes clear that health data should not be stored outside Kenyan territory.

Health data protection and privacy regulations are now well established in Kenya, and therefore so are the governance demands for challenge participants. Organisational capacity to meet these provisions in Kenya should be considered given the very recent introduction of the Data Protection Act with parts, such as the process for approval of data transfer safeguards, still in implementation. Low data management capacity and skills at primary care facilities, as well as fragmentation in the health system are also cited as a constraint on data use. Professional legal support, and ensuring sufficient time to address legal and capacity issues, should be budgeted for by challenge participants to access and use personal health data.

In terms of municipal services, the National Solid Waste Management Strategy (2014)\textsuperscript{213} by the National Environment Management Authority (NEMA) is the most recent government action establishing a common platform for action between stakeholders to systematically improve waste management. It introduces a new approach for improved waste management in Kenya to create wealth and employment and reduce pollution of the environment. However, the county government lacks operational capacity in both staff and equipment to undertake SWM in the city. The regulator NEMA is equally challenged and cannot effectively supervise SWM in the city.

Although digital literacy and data skills are a focus of government initiatives, health information system investment still remains low. Sustainability funding and end user engagement will support uptake and implementation of the tools designed in the challenge and may present opportunities for further funding within the Data for Science and Health programme by building scalable models across the country, or models that could be adapted for use in other areas of public health.

Full legal, policy and regulatory review for Kenya can be found in Appendix 17.

Nigeria

Generally, the provision of healthcare services in Nigeria is largely borne by the three levels of government – Federal government, State government and Local government, with a substantial level of involvement by the private sector. The key stakeholders organisations involved in the provision of healthcare in Nigeria are the Federal Ministry of Health (“FMoH”) and its agencies (including, among others, the National Health Insurance Scheme, the Nigerian Institute of Medical Research, the National Primary Health Care Development Agency, the National Agency for Food and Drug Administration and Control, National Agency for the Control of AIDS, the Ministry of Health in each State, the Local Government authorities, and private healthcare providers. This multi-level governance structure means that during the setup phase it will be vital to engage with stakeholders from all levels and through that be very clear about whether the challenge is running on a national, state or city level. This is likely to be determined by the level of engagement from each group.


\textsuperscript{213} National Environment Management Authority (2014), National Solid Waste Management Strategy, Nairobi.
Data protection in the health sector is mainly regulated by the National Health Act and the Nigeria Data Protection Regulation, 2019 (“NDPR”) issued by the National Information Technology Development Agency (“NITDA”) (the body authorised by the National Information Technology Development Agency Act, 2007 to develop guidelines in relation to the exchange of electronic data and information and other forms of electronic communication transactions). The National Health Act contains provisions that require service providers to maintain confidentiality in handling patient’s information, restricts the disclosure of information and requires service providers to have control measures for preventing unauthorised access to the information of a patient. The FMOH is the government parastatal empowered to regulate the health system in Nigeria and formulate standards for rendering health services in Nigeria, including ensuring and promoting compliance with the confidentiality and privacy provisions of the law. They would therefore need to be a key stakeholder in this challenge in order to ensure that the health angle was taken fully into consideration.

The NDPR applies to all transactions intended for the processing of personal data of natural persons residing in Nigeria or residing outside Nigeria who are Nigerian citizens. Personal data is defined as the information relating to an identified or identifiable person, as such, it applies to the processing of personal data in Nigeria. Therefore, the processing of personal data within the health sector must conform with the provisions of the NDPR in addition to the provisions of the National Health Act. For the challenge this means that any data which can be tied to an individual, be it through municipal records, health data or other sources, usage must adhere to the NDPR and legal support will be required to ensure proper adherence for all participants. However, in comparison to other challenges, there is likely to be a lower amount of personal data being handled so the risk is relatively low.

The Personal Information and Data Protection Bill, 2017 (“PIDPB”) proposes certain changes to privacy or data protection laws in Nigeria that may affect the use and sharing of Data. If this bill is passed into law, it will serve as the principal legislation on data protection in Nigeria. However, the timeline for the passing of the bill is dependent on how quickly the National Assembly is able to pass it for presidential assent. Should it pass prior to the start of the challenge, it is vital to consider the knock on impacts on the capacity of data stewards, participants and beneficiaries to comply with new legislation and therefore additional support will be required.

With regards to intellectual property protection available for health data, health data may be protected as a copyright, trade secret or patent. However, the first test is to ensure that such data meets the various eligibility requirements under Nigerian law. Currently, Nigerian intellectual property laws afford insufficient protection for those who invest time, effort, and money in collecting, compiling and sharing data so there is an increasing reliance on contracts. There are however no proposed changes to the existing intellectual property legislation in Nigeria that may affect the use and sharing of Data. This legal grey area will need to be taken into consideration early in the set up process when considering the IP of the outputs to ensure equitable benefit across the participants, data stewards and stakeholders.

Full legal, policy and regulatory review for Kenya can be found in Appendix 21.
Data Inventory

Below is a list of the data sources that would be relevant for this challenge. Below that are excerpts from research demonstrating the relevant urban health indicators and potential data sources which will correlate to them.

<table>
<thead>
<tr>
<th>Data type</th>
<th>Organisation/ Source</th>
<th>Availability</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>International Environmental/ Spatial/ Remote Sensing Data Sources</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Environmental and population data</td>
<td>UN GRID Core Data Sets Portal</td>
<td>Openly available</td>
<td>biosphere, cyclone, Disaster Risk Reduction, Biodiversity, Climate Change data</td>
</tr>
<tr>
<td>Environmental data</td>
<td>the WorldClim2 database</td>
<td>Openly available</td>
<td>A database of high spatial resolution global weather and climate data. These data can be used for mapping and spatial modeling. The data are provided for use in research and related activities; You can download gridded weather and climate data for historical (near current) and future conditions.</td>
</tr>
<tr>
<td>EO data</td>
<td>CCI Africa Land Cover map (20m resolution)</td>
<td>Openly available</td>
<td>A prototype high resolution Land Cover map over Africa at 20m, based on 1 year of Sentinel-2 A satellite observations from December 2015 to December 2016. The Coordinate Reference System used for the global land cover database is a geographic coordinate system (GCS) based on the World Geodetic System 84 (WGS84) reference ellipsoid.</td>
</tr>
<tr>
<td>EO data</td>
<td>Global Earth Observation System of Systems (GEOSS)</td>
<td>Openly available</td>
<td>GEOSS is a set of coordinated, independent Earth observation, information and processing systems that interact and provide access to diverse information for a broad range of users in both public and private sectors. The ‘GEOSS Portal’ offers a single Internet access point for users seeking data, imagery and analytical software packages relevant to all parts of the globe.</td>
</tr>
<tr>
<td>EO data</td>
<td>African Regional Open Data Cube</td>
<td>Openly available</td>
<td>The Africa Regional Data Cube (ARDC) is a tool that provides access to the latest Earth observation data and satellite technology. It is in the process of being expanded to cover all of Africa.</td>
</tr>
<tr>
<td>EO data</td>
<td>Nasa Earth Observation Data</td>
<td>Openly available</td>
<td>A range of openly available satellite data is available through this portal including near real-time data through the Land, Atmosphere Near real-time Capability for EOS (LANCE).</td>
</tr>
<tr>
<td>Data Type</td>
<td>Source</td>
<td>Availability</td>
<td>Description</td>
</tr>
<tr>
<td>----------------------------</td>
<td>---------------------------------------------</td>
<td>--------------------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>EO data</td>
<td>US Landsat Data</td>
<td>Openly available</td>
<td>Portal which provides links to access points of landsat data through visualisations (e.g: EarthExplorer: Graphical interface used to define areas of interest by address, zip code, place name, or using the map) or as bulk data access options.</td>
</tr>
<tr>
<td>EO data</td>
<td>Copernicus EO Data</td>
<td>Openly available</td>
<td>Copernicus builds on a constellations of satellites making millions an impressive number of of daily observations, as well as on a global network of thousands of land-, air- and marine-based sensors to create the most detailed pictures of Earth.</td>
</tr>
<tr>
<td>Geospatial data</td>
<td>Open Street Map</td>
<td>Openly available</td>
<td>OpenStreetMap is a map of the world, created by people like you and free to use under an open license.</td>
</tr>
<tr>
<td>Geospatial data</td>
<td>UAV Data</td>
<td>To be sourced/</td>
<td>EO data may not be of sufficient resolution to see specific indicators relating to public health and therefore UAV data may be required to add greater detail to specific areas.</td>
</tr>
<tr>
<td>Municipal data</td>
<td>the Africa Electricity Grids Explorer</td>
<td>Openly available</td>
<td>An up to date collection of open data on grid networks in Africa and the Middle East.</td>
</tr>
<tr>
<td>Socio-economic indicators</td>
<td>Open Data for Africa</td>
<td>Openly available</td>
<td>The AfDB Statistical Data Portal has been developed in response to the increasing demand for statistical data and socio-economic indicators relating to African Countries. The Portal provides multiple customized tools to gather indicators, analyze them, and export them into multiple formats.</td>
</tr>
<tr>
<td>Population data</td>
<td>European Commission: Global Human Settlement City Model</td>
<td>Openly available</td>
<td>The Urban Centre Database GHS-UCDB R2019A describes more than 10,000 urban centres identified by the application of the &quot;Degree of Urbanization&quot; model to the GHSL baseline data. This dataset is a successful result of the scientific cooperation among the partners of the GEO Human Planet Initiative.</td>
</tr>
<tr>
<td>High Resolution Population Data Maps</td>
<td>Humanitarian Data Exchange</td>
<td>Openly available</td>
<td>The population of the world, allocated to 1 arcsecond blocks.</td>
</tr>
<tr>
<td>Administrative Boundary Data</td>
<td>GADM</td>
<td>Openly available</td>
<td>Maps and spatial data for all countries and their subdivisions. You can browse maps or download the data to make custom maps.</td>
</tr>
<tr>
<td>Administrative Boundary Data</td>
<td>DIVA</td>
<td>Openly available</td>
<td>International administrative boundaries, inland water, roads, railroads, elevation, land cover, place names and coordinates.</td>
</tr>
</tbody>
</table>

**International Health Data Sources**

<table>
<thead>
<tr>
<th>Data Type</th>
<th>Source</th>
<th>Availability</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Health facility data</td>
<td>Sub-Saharan Public Hospitals Geo-coded database</td>
<td>Openly available</td>
<td>A geocoded inventory of public hospitals across 48 countries and islands of sub-Saharan Africa from 100 different sources. A cost distance</td>
</tr>
</tbody>
</table>

**Notes:**
- **EarthExplorer**: A graphical interface used to define areas of interest by address, zip code, place name, or using the map.
- **Openly available**: Indicates that the data is freely accessible and usable.
- **To be sourced/gathered**: Indicates that the data is not readily available and may require additional effort to obtain.
- **Openly available**: Indicates that the data is freely accessible and usable.
- **Maps**: Indicates the availability of spatial data in map format.
- **Download data**: Indicates the availability of spatial data for download.
<table>
<thead>
<tr>
<th>Surveillance Data</th>
<th>INDepth Network</th>
<th>TBC</th>
<th>algorithm based on the location of 4908 public hospitals, population distributions and road networks were used to compute the proportion of populations living within a combined walking and motorised travel time of 2 hours to emergency hospital services.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Global Health Systems</td>
<td>World Bank - World Development Indicators</td>
<td>Openly available</td>
<td>A global network of health and demographic surveillance systems (HDSSs) that provide a more complete picture of the health status of communities. Since they collect data from whole communities over extended time periods, they more accurately reflect health and population problems in low- and middle-income countries (LMICs).</td>
</tr>
<tr>
<td>Health Data management Platform</td>
<td>DIHS2</td>
<td>Closed</td>
<td>DHIS2 is the world’s largest health management information system (HMIS) platform, in use by 72 low and middle-income countries. DHIS2 is typically used as national health information systems for data management and analysis purposes, for health program monitoring and evaluation, as facility registries and service availability mapping, for logistics management and for mobile tracking of pregnant mothers in rural communities.</td>
</tr>
<tr>
<td>Infectious Diseases</td>
<td>Infections Diseases Data Observatory</td>
<td>TBC</td>
<td>The Infectious Diseases Data Observatory (IDDO) assembles clinical, laboratory and epidemiological data on a collaborative platform to be shared with the research and humanitarian communities.</td>
</tr>
<tr>
<td>Observational Health Databases</td>
<td>The Observational Health Data Sciences and Informatics (ODHSI)</td>
<td>ODHSI's OMOP Common Data Model allows for the systematic analysis of disparate observational databases.</td>
<td></td>
</tr>
<tr>
<td><strong>Nigeria Specific Data Sources</strong></td>
<td></td>
<td></td>
<td>They could provide a number of highly valuable datasets, including access to the NigeriaSat-2 and Ortho-rectified Spot-5 satellite data. In addition, NARSDA is in the process of taking over the ownership and maintenance of the Grid3 Nigeria georeferenced data infrastructure which includes over 980 publicly available data sets in 37 states relating to administrative boundaries, population, energy, settlements, health, facilities and others.</td>
</tr>
<tr>
<td>National geospatial data + environmental policy data</td>
<td>National Space Research and Development Agency (NASRDA)</td>
<td>Available through a partnership + openly available</td>
<td>- Lagos Waste Management Authority (LAWMA)  - Lagos State Environment Protection Agency</td>
</tr>
<tr>
<td>Service data</td>
<td>City and state authorities</td>
<td>Available through a partnership</td>
<td></td>
</tr>
</tbody>
</table>

147
<table>
<thead>
<tr>
<th>Ground truthing/ community data</th>
<th>Data Science Nigeria/ Slum Dwellers International</th>
<th>To be collected</th>
<th>Both Data Science Nigeria and SDI have experience in community data collection and could be commissioned to collect relevant data.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Demographic and Health Surveys (DHS)</td>
<td>National Population Commission/ USAID</td>
<td>Available but out of date</td>
<td>The most recent survey is from 2018.</td>
</tr>
<tr>
<td>Census</td>
<td>Nigeria Bureau of Statistics</td>
<td>Available but out of date</td>
<td>The most recent survey is from 2006</td>
</tr>
<tr>
<td>Household surveys</td>
<td>Nigeria Bureau of Statistics</td>
<td>Available but out of date</td>
<td>2013 January - 2015 December</td>
</tr>
</tbody>
</table>

*Further data will be available through partnerships with relevant stakeholder groups (slum organisations, local environmental NGOs, public sector waste management and private sector CSR as detailed in the Data Providers and Delivery Partners section)*

### Kenya Specific Data Sources

<table>
<thead>
<tr>
<th>Demographic and Health Surveys (DHS)</th>
<th>Kenyan Government/ USAID/ Worldbank</th>
<th>Available but out of date</th>
<th>The most recent survey is from 2014</th>
</tr>
</thead>
<tbody>
<tr>
<td>Census</td>
<td>Kenya National Bureau of Statistics</td>
<td>Available</td>
<td>Most recent census is from 2019, conducted using mobile technology was used during mapping and enumeration, in adherence to the UN recommendations for the 2020 round of censuses on adoption of use of technology.</td>
</tr>
<tr>
<td>Household surveys</td>
<td>Kenya National Bureau of Statistics</td>
<td>Available</td>
<td>The most recent survey is from 2015/16</td>
</tr>
<tr>
<td>Environmental Policy</td>
<td>National Environment Management Authority (NEMA)</td>
<td>Available through a partnership</td>
<td>This department has oversight over solid waste, pollution and flood management across Kenya. They could supply policy level data.</td>
</tr>
<tr>
<td>City and service level data</td>
<td>State and City Authorities</td>
<td>Available through a partnership</td>
<td>NEMA have the power to enable data sharing through state and city authorities for environmental services</td>
</tr>
<tr>
<td>Health Policy</td>
<td>Ministry of Health and National Bureau of Statistics</td>
<td>Available through a partnership</td>
<td>NEMA have agreed to act as a convener with other key governmental stakeholders in order to facilitate access to data from the Ministry of Health and National Bureau of Statistics.</td>
</tr>
<tr>
<td>Urban Health surveillance data</td>
<td>Nairobi Urban Health and Demographic Surveillance System (NUHDSS)</td>
<td>Openly available</td>
<td>The system collects data in two slum communities in Nairobi (Korogocho and Viwandani) on demographic events (births, deaths, and migrations), health outcomes (morbidity, cause of death through verbal autopsy, child vaccination, and nutrition) and socio-economic outcomes (marriage, education, livelihood, and housing characteristics).</td>
</tr>
<tr>
<td>Deprivation mapping</td>
<td>SLUMAP</td>
<td>TBC</td>
<td></td>
</tr>
<tr>
<td>---------------------</td>
<td>--------</td>
<td>-----</td>
<td></td>
</tr>
<tr>
<td></td>
<td>An extensive research project looking at slum mapping in Kenya through remote sensing which could provide an excellent data and conceptual framework to be utilised for public health purposes. They will have delivered the research by 2021.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Ground Truthing/Community Data</th>
<th>Map Kibera / Slum Dwellers International</th>
<th>Openly available/ to be collected</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Map Kibera is a very established community mapping initiative which works across Nairobi to support slum mapping based on Open Street Map. Slum Dwellers International has a Kenyan Federation which also does annual data collection on slum communities across the country. Both organisations have been engaged with and are interested in partnering.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Field data</td>
<td>EO</td>
</tr>
<tr>
<td>-------------------------------------------------------------------------------------------</td>
<td>------------</td>
<td>----</td>
</tr>
<tr>
<td><strong>Slum area model training data</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Area classification during census/survey</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Participatory slum mapping</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Geotagged photos (e.g., Flickr)</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Online crowdsourced mapping</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Manually digitize satellite image</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Govt-registered slum locations</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Social/environmental risk</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Climate (precipitation, temperature)</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Green space type, coverage</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Hazardous location—flood zone, slope</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Median household/percapita income</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Mobile phone use (e.g., number calls)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mobile phone top-up (e.g., amount)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mobile phone mobility patterns</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mobile phone social network metrics</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Open space coverage</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Percent HHs nondurable floor, roof, wall</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percent HHs overcrowding</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percent HHs unimproved sanitation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percent HHs unimproved water</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proximity, travel time to CBD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proximity to landcover type (e.g., marsh)</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Proximity to high-voltage power lines</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proximity to highways, major roads</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proximity to railway</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proximity to river, stagnant water body</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td><strong>Lack of facilities/infrastructure</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nighttime light intensity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Open drains present</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Proximity, density health facilities</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proximity, density schools</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proximity to public transport stop/line</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Road coverage</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Road material (e.g., paved)</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Road pattern</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Road repair conditions</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Road width/height (e.g., local, main)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Unplanned urbanization</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Building coverage, density</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Building height, shadow</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Building organization</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Building roof material, color</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Building footprint (size, shape)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Assessment of technical feasibility, resources, and source data needed to generate area-level health determinant indicators in LMICs, by Bellagio slum area definition domain \(^{215}\)

<table>
<thead>
<tr>
<th>Neighbourhood-level health determinant</th>
<th>Technical feasibility</th>
<th>Resources needed</th>
<th>Data availability</th>
<th>Main data source(s)</th>
<th>Optional or supplemental data source(s)</th>
<th>Existing products (resolution)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Social/environmental risk</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Civil engagement (e.g. voting rate)</td>
<td></td>
<td></td>
<td></td>
<td>Voting record count</td>
<td>Gridded population model Census</td>
<td>HDX (^{226})</td>
</tr>
<tr>
<td>Climate - ave monthly/annual precipitation</td>
<td></td>
<td></td>
<td></td>
<td>LR imagery</td>
<td>WorldClimate v2 (^{227}), IRI Climate Library (^{228})</td>
<td></td>
</tr>
<tr>
<td>Climate - ave monthly/annual temperature</td>
<td></td>
<td></td>
<td></td>
<td>LR imagery</td>
<td>WorldClimate v2 (^{227}), IRI Climate Library (^{228})</td>
<td></td>
</tr>
<tr>
<td>Green space boundaries</td>
<td></td>
<td></td>
<td></td>
<td>GIS + HR imagery</td>
<td>GIS + VHR imagery</td>
<td></td>
</tr>
<tr>
<td>Green space boundaries, by type (e.g. maintained, overgrown, garden)</td>
<td></td>
<td></td>
<td></td>
<td>GIS + VHR imagery</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Open space boundaries</td>
<td></td>
<td></td>
<td></td>
<td>VHR imagery</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Flood zone</td>
<td></td>
<td></td>
<td></td>
<td>HR imagery + Terrain model VHR imagery + Terrain model</td>
<td>Soil maps</td>
<td>SRTM (90m)(^{229})</td>
</tr>
<tr>
<td>Steep slopes</td>
<td></td>
<td></td>
<td></td>
<td>HR imagery + Terrain model VHR imagery + Terrain model</td>
<td>SRTM (90m)(^{229})</td>
<td></td>
</tr>
<tr>
<td>Land use change</td>
<td></td>
<td></td>
<td></td>
<td>GlobalLand30 + GIS + Fonte, et al. 2017 model (^{230})</td>
<td>OpenStreetMap(^{231}), CCI Land Cover (20m)(^{232}), GlobalLand30 (30 m)(^{233})</td>
<td></td>
</tr>
<tr>
<td>Proximity to landcover type (e.g. marsh, water)</td>
<td></td>
<td></td>
<td></td>
<td>GIS data</td>
<td>HR imagery</td>
<td>OpenStreetMap(^{231}), CCI Land Cover (20m)(^{232}), GlobalLand30 (30 m)(^{233})</td>
</tr>
<tr>
<td>Household/per-capita income</td>
<td></td>
<td></td>
<td></td>
<td>Survey Census</td>
<td>HR imagery + GIS data</td>
<td></td>
</tr>
<tr>
<td>Percent households nondurable floor, roof, wall</td>
<td></td>
<td></td>
<td></td>
<td>Survey Census</td>
<td>HR imagery + GIS data</td>
<td></td>
</tr>
<tr>
<td>Percent households overcrowding</td>
<td></td>
<td></td>
<td></td>
<td>Survey Census</td>
<td>HR imagery + GIS data</td>
<td></td>
</tr>
<tr>
<td>Percent households unimproved sanitation</td>
<td></td>
<td></td>
<td></td>
<td>Survey Census</td>
<td>HR imagery + GIS data</td>
<td>DHS (5km)(^{236})</td>
</tr>
<tr>
<td>Percent households unimproved water</td>
<td></td>
<td></td>
<td></td>
<td>Survey Census</td>
<td>HR imagery + GIS data</td>
<td>DHS (5km)(^{236})</td>
</tr>
<tr>
<td>Mobile phone use (e.g. number calls)</td>
<td></td>
<td></td>
<td></td>
<td>CDR</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mobile phone top-up (e.g. amount)</td>
<td></td>
<td></td>
<td></td>
<td>CDR</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mobile phone mobility patterns</td>
<td></td>
<td></td>
<td></td>
<td>CDR</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mobile phone social network metrics</td>
<td></td>
<td></td>
<td></td>
<td>CDR</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proximity to high voltage power lines</td>
<td></td>
<td></td>
<td></td>
<td>GIS data</td>
<td>Africa Grid(^{234}), OpenStreetMap(^{231})</td>
<td></td>
</tr>
<tr>
<td>Proximity to highways, major roads</td>
<td></td>
<td></td>
<td></td>
<td>GIS data</td>
<td>HDX(^{20}), DivA(^{20}), OpenStreetMap(^{231})</td>
<td></td>
</tr>
<tr>
<td>Proximity to railway</td>
<td></td>
<td></td>
<td></td>
<td>GIS data</td>
<td>HDX(^{20}), DivA(^{20}), OpenStreetMap(^{231})</td>
<td></td>
</tr>
<tr>
<td>Proximity to river</td>
<td></td>
<td></td>
<td></td>
<td>GIS data</td>
<td>OpenStreetMap(^{231})</td>
<td></td>
</tr>
<tr>
<td>Proximity to central business district</td>
<td></td>
<td></td>
<td></td>
<td>GIS data</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Travel time to central business district</td>
<td></td>
<td></td>
<td></td>
<td>GIS data + cost/distance model</td>
<td>VHR imagery</td>
<td></td>
</tr>
<tr>
<td>Crime rates</td>
<td></td>
<td></td>
<td></td>
<td>Crime reports</td>
<td>Gridded population model Census</td>
<td></td>
</tr>
<tr>
<td>Lack of facilities / infrastructure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Businesses number, type</td>
<td></td>
<td></td>
<td></td>
<td>GIS data</td>
<td></td>
<td>OpenStreetMap(^{231})</td>
</tr>
<tr>
<td>Solid waste system coverage</td>
<td></td>
<td></td>
<td></td>
<td>City planning documents</td>
<td>Gridded population model Census</td>
<td></td>
</tr>
<tr>
<td>Energy, telecom coverage</td>
<td></td>
<td></td>
<td></td>
<td>City planning documents</td>
<td>Gridded population model Census</td>
<td></td>
</tr>
<tr>
<td>Number of intersections</td>
<td></td>
<td></td>
<td></td>
<td>GIS data</td>
<td></td>
<td>HDX(^{20}), OpenStreetMap(^{231})</td>
</tr>
</tbody>
</table>
Appendices

The following appendices can be found in the supporting zip file:

Appendix 1: Prize Programme + Prize Specific Theory of Change
Appendix 2: Prize Programme Budget
Appendix 3: Prize Programme + Prize Specific Risk Register
Appendix 4: Prize Programme Communications Strategy
Appendix 5: List of all engaged stakeholders across scoping phase
Appendix 6: Data Science Challenge Prizes for Health: A Playbook
Appendix 7: Data Science Challenge Prize Longlist
Appendix 8: Data Science Challenge Prize Longlist Scoring Sheet
Appendix 9: Benchmarking of health tech innovation prizes
Appendix 10: Programme and Prize Timelines
Appendix 11: Mental health data ecosystem map
Appendix 12: Snakebite data ecosystem map
Appendix 13: DRI data ecosystem map
Appendix 14: Urban Health data ecosystem map
Appendix 15: UK legal, regulatory and policy assessment
Appendix 16: India legal, regulatory and policy assessment
Appendix 17: Legal, legal, regulatory and policy assessment for Kenya
Appendix 18: Legal, legal, regulatory and policy assessment for Uganda
Appendix 19: Legal, legal, regulatory and policy assessment for Malawi
Appendix 20: Legal, regulatory and policy assessment for Nigeria