

## HEALTH DATA FOR LONDON STRATEGY

### SUMMARY

**Our vision** is to use the power of data to drive improvements in health and care for Londoners, and to propel London to be a global powerhouse for AI and digital therapeutic innovation, and diagnostics development.

London has many clinical and academic data strengths, and a thriving technology and life sciences innovation cluster. But the London system is still too fragmented. Without access to high-quality, system-wide health data in near real time, it is not possible to undertake the research, analysis and innovation needed to tackle the health inequalities in London.

We will work in partnership to safely join up data from across London and drive collaboration between existing initiatives to make London a world-leader in the use of data to improve health outcomes; to provide insights and intelligence and connect research and clinical care to create a genuinely learning health system.

One London, Health Data Research UK and our partners set out a three-part strategy to enable London to make better use of its rich data assets. A coordinated approach will position London as the global location of choice for academic and industrial innovation with health data, attracting inward investment and delivering rapid benefits to the NHS, patients, and the public.

### INTRODUCTION

In February 2020, London's Institutional health leaders<sup>1</sup> - NHS London, the Greater London Authority and the London Academic Health Science Networks - sponsored a London-wide consultation to explore how London could develop as a city for data science for health and innovation, and to set out options for operational arrangements.

While the work was temporarily delayed by COVID-19, it quickly became clear that data was a critical enabler for the NHS to tackle the pandemic. COVID-19 highlighted the need to integrate health and care data to provide information in real-time, and reinforced the importance of developing a broader health data strategy for London. Over the past six months we have consulted with more than fifty stakeholders across London, sought international advice and comparison, and worked closely with an Expert Reference Group of 15 academic and NHS leaders to develop this vision for a London Health Data Strategy.

---

<sup>1</sup> The Mayor's Office, NHS London, University College London, Kings College London, Imperial College, Queen Mary University of London, London School of Hygiene and Tropical Medicine

## The challenge

London faces "stark health inequalities". Life expectancy varies significantly across different boroughs in London, and there is huge variation in how long Londoners can expect to live in good health. Too many people are living with multiple and complex co-morbidities and long-term conditions. A woman living in Tower Hamlets may live up to a third of her life in poor health; one in four Londoners will experience poor mental health.<sup>2</sup> London has particularly high levels of childhood obesity, with 2 in 5 children overweight or obese by the age of ten.<sup>3</sup>

Between March and July 2020, London had the highest proportion of deaths involving COVID-19 across the UK.<sup>4</sup> The COVID-19 pandemic has demonstrated the absolute necessity of connecting health, care, and cross-sectoral datasets to provide information about the whole system to improve the response. Dashboards, reporting, and analytics helped inform planning and deliver better quality care across London. However, it also highlighted the significant difficulty of getting high quality data from across multiple sites in real time.

### Case Study 1: Improving childhood immunization in London

***London is consistently the worst performing region in England for childhood immunization, and rates are continuing to decline.*** No London local authority meets the WHO target of 95% vaccination. In 2019, this led to rising cases of measles, with 58% of all UK cases in London.

There is wide variation in immunization rates between CCGs in London, with rates of MMR immunization ranging from 90.9% down to 74.3%. These figures have fallen even further as a result of COVID. In east London MMR timeliness fell from 71.4% in first 6 months of 2019 to 66.1% in equivalent time period of 2020. Vaccination rates are worst in children from most deprived areas and for those from Black and Asian backgrounds.

Access to real-time, structured high-quality data could transform the provision of immunization. A Cochrane review has shown that patient reminder and recall systems, in primary care settings, are likely to be effective at improving immunisation uptake, and a systematic review has suggested that locally designed, multicomponent interventions can be effective in ethnically diverse, deprived populations. Providing access to high-quality data and risk stratification tools would enable local delivery teams to design approaches to ensure all children in London are adequately protected.

*[From: C. Dezateux, Digital innovation for equitable delivery of timely childhood immunisations, presentation to London Data Strategy ERG meeting]*

---

<sup>2</sup> <https://www.healthylondon.org/wp-content/uploads/2018/01/London-Health-Commission-Global-City-Comparisons-overview.pdf>

<sup>3</sup> The rate of childhood obesity in London is 37.9% [London Datastore – Greater London Authority](#)

<sup>4</sup> <https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/bulletins/deathsinvolvingcovid19bylocalareasanddeprivation/deathsoccurringbetween1marchand31may2020>

The Mayor's vision for London set a shared ambition to make London "the healthiest global city, and the best global city in which to receive health and care services".<sup>5</sup> Without a joined-up health data strategy, enabling insights and intelligence to be shared across the system, this will not be possible. The case studies (see boxes) illustrate current challenges, including poor childhood immunisation, low cancer screening rates, and the need to manage cardiovascular risks, that can only be tackled with better access to data. London is currently lacking system-wide intelligence, and fragmentation prevents the delivery of care being informed by research and analysis. Without high-quality joined-up health data, it will not be possible to tackle the health challenges facing Londoners, or to improve the health of the city.

## The opportunity

London has some of the richest health care and clinical research data assets world-wide. **The OneLondon Local Health & Care Record Exemplar** has made great strides towards linking up health information so that there is a genuine single health record for everyone, with professionals able to see all relevant information to provide safer and more coordinated care. **The London Care Record**<sup>6</sup> currently covers a population of 6.5m Londoners and will, **by March 2021, include the full population of 9 million citizens**, the first time information of this richness and scale has been assembled together.

But data alone is not enough. The ability to analyse and use it effectively, to extract useful knowledge, is crucial. London can draw on world class strengths across three areas to achieve this:

- **Clinical expertise:** London has 1400 General Practices, 34 Acute and Foundation Trusts, 14 Teaching Hospitals, 32 Clinical Commissioning Groups, 5 STPs, 3 Mental Health Trusts and 2 Community Trusts. While in the past this fragmentation has been a challenge, the introduction of **five Integrated NHS Care Systems** gives the potential to coordinate across London much more effectively.
- **Academic strength:** London hosts three outstanding **Academic Health Science Centres** and three **Academic Health Science Networks** all with international recognition. The HDR-UK London research site brings together all five of London's major universities in biomedicine and health working in partnership to address interconnected research priorities. London hosts three NIHR Applied Research Centres in North Thames, North West London and South London.<sup>7</sup>
- **An industrial hub:** London has powered itself to be world-leading in data science and artificial intelligence (AI) research and innovation; London is arguably Europe's data capital. **Tech Nation**,<sup>8</sup> the growth platform for tech companies and leaders, MedCity<sup>9</sup> the cluster organisation for the world-leading health and life sciences sector in London and the Greater South East of England and the Alan Turing Institute,<sup>10</sup> the UK Institute for data science, are all hosted in London. Four of

---

<sup>5</sup> *Our Vision for London* (2019) was developed in collaboration by the NHS in London, Public Health England, Health Education England, London Councils (representing London's boroughs), local borough partnerships, and the Mayor of London  
<https://www.england.nhs.uk/london/wp-content/uploads/sites/8/2019/10/London-Vision-2019-FULL-VERSION-1.pdf>

<sup>6</sup> <https://www.onelondon.online/>

<sup>7</sup> <https://www.nihr.ac.uk/explore-nihr/support/collaborating-in-applied-health-research.htm#:~:text=NIHR%20Applied%20Research%20Collaborations&text=The%20ARCs%2C%20announced%20in%20July,system%20both%20locally%20and%20nationally.>

<sup>8</sup> <https://technation.io/>

<sup>9</sup> <https://www.medcityhq.com/>

<sup>10</sup> <https://www.turing.ac.uk/>

the seven *Health Data Research Hubs*<sup>11</sup>, and the *London Medical Imaging and Artificial Intelligence Centre for Value-Based Healthcare*,<sup>12</sup> all funded as part of the Industry Strategy Challenge Fund, have London leadership. The technology industry in London and the South East was worth £47 billion in Gross Value Added (GVA) in 2019. Demonstrated by investments into infrastructure, particularly around Kings Cross with the arrival of GSK AI and MSD, the opportunities to leverage this and attract more inward investment are immense. But this will require us to streamline connectivity and visibility of data assets.

Despite these individual strengths, London has not yet been able to harness them effectively together. The lack of a coherent approach means that London is missing the opportunity to attract inward investment. Without rapid action to address fragmentation, London risks losing its competitive advantage. A single strategy, linking the capabilities of the NHS in London with London's global Universities, together with the leadership and support of the Mayor's Office and a joint strategy with MedCity, is crucial to reap the health and economic benefits of data driven innovation in healthcare, and to improve the health of Londoners.

### A mandate from Londoners

There is a clear mandate from Londoners, expressed through the *Conversation with Londoners* deliberative event (see Annex B), that data should be consistently joined-up across London to support individual care, proactive care, planning and research. Making better use of de-personalised data for planning health services was considered a "must-do" by this group of Londoners. There is a clear expectation that data should be consistently joined-up with a single approach across all London health and care organisations. We owe it to patients to achieve this.

#### Case Study 2: Targeted cancer screening for London

Bowel cancer is the second most common cause of cancer death in the UK. If diagnosed early, nearly everyone with bowel cancer can be successfully treated. However, rates of survival drop significantly the later the presentation.

Early detection is therefore crucial, but ***boroughs in London have some of the lowest rates of screening for bowel cancer in age 60 – 69 population***, and the lowest percentage of cancers receiving first definitive treatment within two months of referral from GP. London has some of the highest rates for emergency presentation for bowel cancer and the highest spend on non-elective admissions for cancer. Deaths from cancer are more common in socioeconomically deprived areas.

This situation has been exacerbated by COVID-19, with almost all cancer screening initially suspended. BAME patients have been particularly unwilling to attend sites during the pandemic. There is an urgent need to understand the problem, and to assess inequalities and deprivation impacts by CCG. Access to real world evidence would enable targeted information to identify those patients which require immediate prioritisation in order to improve outcomes.

[From: C Davie, DataCan Accelerator Project proposal]

<sup>11</sup> <https://www.hdruk.ac.uk/helping-with-health-data/health-data-research-hubs/>

<sup>12</sup> <https://www.kcl.ac.uk/bmeis/research-impact/london-medical-imaging-and-ai-centre-for-value-based-healthcare>

## OUR VISION

Our vision is to use the power of data to drive improvements in health and care for Londoners. We will join up data from across the London system to:

- Provide insights and intelligence;
- Connect research and clinical care to create a genuinely learning health system;
- Drive collaboration between existing initiatives to make London a world-leader in the use of data to improve health outcomes.
- Propel London to be a powerhouse for AI and digital therapeutic innovation, and diagnostics development with underpinning manufacture and commercialisation capability.

NHS London will work in partnership with the public, London's leading research universities – Imperial College, Kings College, LSHTM, QMUL, UCL – and industry to create a 'one-stop' service for trustworthy, multi-stakeholder utilisation of curated data for public, private and third sector benefit.

The intention is not to join up datasets to create a single vast data store. The aim is to be able to bring together the right data at the right time to address key questions. In this way, research will be able to inform the delivery of care, leading to improvements in health at both individual and population levels.

This will place London in a strong position to partner, both across the UK and internationally. We will be able to share learning with other cities and regions in the UK, forming strong partnerships across the country. And we will be in a position to collaborate more effectively internationally, bringing innovation and investment to the UK.

### A local approach within a national framework

We recognise that joining up systems across London is inherently a 'people and culture' challenge, rather than a technical one. Evidence suggests that it will therefore be most effective to work at a local level, where there is a clear understanding of local need and a strong connection with local champions to build trust and ensure buy-in.

However, it is crucial that a local approach should be developed as part of a national framework. The technical solutions must be the same, to ensure interoperability and opportunity to scale. We will work in partnership with national NHS bodies, including **NHS Digital** and **NHSX** to ensure the strategies are closely aligned and connected.

This reflects a key lesson from the COVID-19 pandemic: the most effective response has needed both national and local elements. A national framework for testing is fundamental to ensure a system-wide overview and understanding of the situation; but local implementation is required to ensure the response is tailored and appropriate in a specific context. Similarly, to improve London mental health services, an understanding of the local context will be essential, and to improve vaccination rates across London, the solutions will need to be applied on the basis of local intelligence (see Case Study 1).

A health system that is genuinely learning and improving has never been effectively achieved in the UK. London has the assets and the ecosystem to make it happen. If London can demonstrate that it can be done and done well, it will then be possible to share learning across the country and to build partnerships with other regions to scale the approach.

### **Case study 3: Tackling cardiovascular disease in London**

Cardiovascular disease (CVD) accounts for a quarter of the life expectancy gap between the rich and poor. It is responsible for 25% of premature mortality and yet is also highly preventable. Across London, late diagnosis and suboptimal treatment of high-risk conditions is common and there is substantial geographical variation. NHS Long-Term Plan modelling suggests that 20,000 strokes and heart attacks could be prevented in London over ten years by optimising the diagnosis and management of high-risk conditions for CVD, including atrial fibrillation, blood pressure and high cholesterol.

Access to comprehensive real-time data could drive understanding, highlight gaps and improve care. Most importantly, local analysis would enable the identification of individual patients who need better care, and allow delivery of focused interventions. There is also the opportunity to link with MedCity's work on fast-tracking diagnostics across cancer/cardiovascular and other priority areas.

*[From M. Kearney, An exemplar for CVD prevention in London presentation]*

### **Why now?**

Economic recovery post-COVID will require an agile coordinated response between NHS, academia industry and activity at scale, so that the opportunities of ARPA, the Comprehensive Spending Review and fiscal policy post-Brexit can be seized. London's data assets must be connected seamlessly to attract inward investment.

We will build on the momentum that has begun with the data-driven response to COVID. The UK Health Data Research Innovation Gateway has ensured the priority datasets for COVID-19 research are findable, accessible, inter-operable and reusable (FAIR) as a single "shop window".<sup>13</sup> This has led to 1,460 researchers coming together delivering 1,000 research outputs demonstrating the value of using health data for rapid response research, to understand more about the disease, map the spread of the virus and test new treatments. For example, research using data about inner city residents from two South London hospitals found that people from Black, South Asian and minority ethnic groups were more likely to get COVID-19 and at greater risk of worse outcomes.<sup>14</sup> Under the convenorship of Health Data Research UK (HDR UK), the priority COVID-19 datasets have been mapped and made discoverable through the Health Data Research Innovation Gateway to allow research which can inform policy and operational decision across the UK. This includes COVID-related data in the Discover-NOW Health Data Research Hub which currently supports access to one of the largest depersonalised linked health data sets in Europe, bringing together real-world patient

<sup>13</sup> <https://www.hdruk.ac.uk/wp-content/uploads/2020/11/201110-Health-Data-Research-UK-COVID-19-fortnightly-update-paper.pdf>

<sup>14</sup> <https://www.hdruk.ac.uk/news/does-your-ethnic-background-put-you-more-at-risk-from-covid-19/>

health data, in a safe, secure Trusted Research Environment for a population of more than 2.3 million in North West London (see Case Study 4).<sup>15</sup>

International exemplars, including the Mayo Clinical Data Analytics Platform (introduced in a three month timeframe on Mayo Clinic Cloud (MCC) hosted by Google) and proposals for the Nordic Commons<sup>16</sup> demonstrate what can be achieved (see Annex F). But they also highlight how London will lose out if it is unable to join up activities and compete effectively. London is in prime position to deliver impact aligned to the NHS Long Term Plan<sup>17</sup>, the Life Science Industrial Strategy<sup>18</sup> and the goals of the expanded Accelerated Access Collaborative (AAC)<sup>19</sup> if we act now.

#### **Case study 4: Responding to COVID-19**

The Discover-NOW Trusted Research Environment has over **30 research and innovation projects** underway using de-identified Real World Evidence. For example, they have identified people at risk of COVID-19 and used targeted interventions to reduce their risk.

Using a data-driven approach to work across local healthcare and government, they have been able to link primary care, local authority, and testing data with geospatial analysis to identify patients and populations suitable for interventions. This has enabled more efficient population health management, and the use of tailored, effective interventions.

*[From B.Klaber et al, Whole Systems Integrated Care & Discover Now presentation]*

***Our proposal is not solely an NHS or an academic venture. No single organisation or sector can deliver the sustainable and broad health and economic benefits that we seek. Effective cross-sectoral partnerships will be crucial. The Delivery Plan outlines an ambitious and agile approach for success.***

## **DELIVERY PLAN**

We recommend three connected workstreams to join up data from across the London system and to drive collaboration between existing initiatives.

### **1. Leadership and coordination**

Strong leadership will be crucial. The first priority will be to establish a high-level policy Board with stakeholders from the London health, research and innovation sectors, together with national NHS representation. This group will be responsible for developing a detailed business plan, ensuring coordination and buy-in across different organisations, establishing specific Task-and-Finish groups to address specific issues, and embedding public involvement from the outset. A key priority for the group will be to secure sustainable funding, with the aim, after initial priming, to be to develop a sustainable, commercial proposition. There is no immediate need to create a new legal entity, but we envisage a formal **Collaboration Agreement** will be required between the NHS (regional and national bodies), London Universities and potential Innovation partners such as MedCity. HDR UK will continue to support as a partner and enabler.

<sup>15</sup> <https://imperialcollegehealthpartners.com/discover-now/>

<sup>16</sup> <https://www.nordforsk.org/2020/vision-nordic-secure-digital-infrastructure-health-data-nordic-commons>

<sup>17</sup> <https://www.longtermplan.nhs.uk/>

<sup>18</sup> [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/650447/LifeSciencesIndustrialStrategy\\_acc2.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/650447/LifeSciencesIndustrialStrategy_acc2.pdf)

<sup>19</sup> <https://www.england.nhs.uk/aac/>

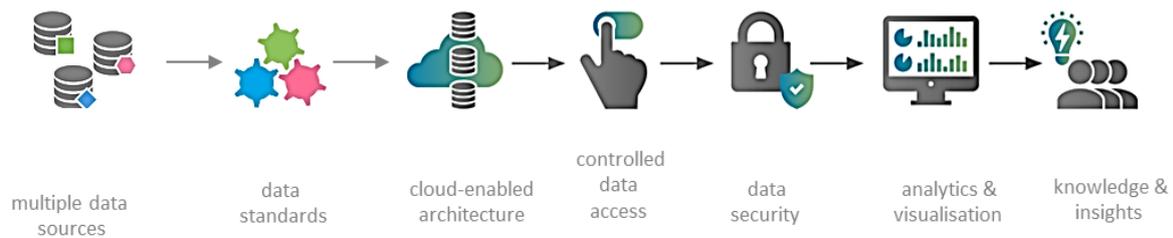
## 2. Sponsoring a portfolio of Health Improvement and Innovation Demonstrator Projects

The Board will select up to five Demonstrator Projects through a fair and transparent process to address London's health priorities, and show how better use of data and innovation can lead to actionable insight. A call for detailed applications will be held in the second phase of delivery, with partnership working as a key criteria. As well as delivering measurable improvements in health, these projects will act as a proof of concept, helping to design, develop and test the systems that will be needed to link up data more broadly across London to systematically embed action at the point of care. Projects are expected to tackle London wide challenges including vaccinations, cancer, cardiovascular disease, childhood obesity and COVID-19.

## 3. Establishing a shared framework for a London Health Data Hub

A key deliverable will be to define, and then implement, the open foundations of a London Health Data hub, with agreement over the harmonisation and implementation of a technology strategy, metadata, standards and interoperability. This will be designed in close partnership with national bodies. The main elements (described in more detail in Annex C) will be:

- **To make data assets visible:** London's datasets, which will be granular, multi-dimensional and wide-ranging (e.g., including unstructured data and citizen-generated data). Priority datasets will be selected to demonstrate the richness of London's data assets, and of value to NHS, industry, and investors. These will be made discoverable on the UK Health Innovation Gateway, ensuring the data is Findable, Accessible, Interoperable and Reusable (FAIR), through shared metadata and advanced discoverability tools.
- **To use Trusted Research Environments (TRE) for data analysis:** Data will be accessible for analysis through secure TREs, cloud-enabled and delivered to robust industrial standards. London already has a number of TREs (see Annex D). The intention is not to select a single supplier but to build interoperability across these different platforms, to ensure a networked data architecture. Over time, the ambition should be to move towards federated across between different TREs, rather than moving data around the system. This will allow researchers and innovators to develop, run and maintain their own analytical applications in a fully-governed, ethical and trustworthy environment without the complexity, cost, or delay of setting up individual systems.
- **To provide expertise for data curation:** Data must be high-quality, reliable and fit-for-purpose. Agreeing standards for data collection across London will improve data quality, ensure consistency and timeliness of data capture and extraction, allow integration of systems, and reduce the costs of interoperability. Dedicated resource for data curation and standardisation will therefore need to be a key element of the framework, working in partnership with the nation-wide approach to data curation. London will also be able to test ways to ensure professional buy-in and institutional commitment to improve data collection. A plan for cross-sector collaboration and skills development on the technology architecture, data wrangling and analytics will be developed with universities, technology start-ups, and the fintech industry.
- **To ensure consistent rules about access:** in line with the recommendations of the Citizens' Summit, governance should be based on the 'five safes' framework. A streamlined approach, with citizen's embedded in decision-making, will be essential to address the complexity of current information governance, and avoid a cautious and risk-averse approach. Above all, London must demonstrate trustworthiness in the way data is handled, ensuring clear public benefit and transparency.



These three workstreams will be delivered through a three-phase plan for the first twelve months. The delivery plan at **Annex A** sets out more detail about these phases for implementation during 2021. We envisage the budget for the first year will be £3-5million, including funding for the Demonstrator Projects (c.£0.5m each). Sustainable funding for 3-5 years will then be required.

### Principles of Delivery

- **Public engagement and involvement:** a commitment to build on the One London citizen's Conversation with Londoners, and put citizens at the heart of decisions.
- **Leveraging assets for London and UK-wide Benefit:** the collaboration will bring together the synergistic strengths identified in London and foster closer working with data scientists and industry to ensure optimal benefits for patients, the NHS and the economy.
- **Interoperability:** London will be known globally for the ability to work across institutions and geographies with no additional effort.

**International Outlook:** The programme will be able to link with UK-wide developments and will align with international efforts.

### The Outcome

This partnership approach will position London as the global location of choice for industrial and academic data driven innovation, and will build a strong evidence base to improve diagnosis, detection, treatment and prevention of disease for Londoners. This will facilitate the delivery of more efficient health and social care, allowing NHS London to make better use of existing resource. It will attract inward investment, with a single offer that will make London the go-to place to undertake data-driven health research and innovation. And above all, it will provide the insight, intelligence and action needed to address health inequalities across London, and to make London the healthiest global city.

## ANNEX A: DELIVERY PLAN

	Phase 1: Convening Jan – March 2021	Phase 2: Mobilise April – Sept 2021	Phase 3: Deliver and accelerate Oct – Dec 2021
<b>Promoting leadership and coordination</b>			
<b>Governance</b>	Agree leadership and establish high-level policy Board to shape direction. Enact a collaboration agreement	Task-and-Finish Groups underway	Task-and-Finish groups complete work
	Establish expert Task and Finish groups: - technology and architecture - standards - legislation, ethics and engagement	Work with NHSX to ensure fair partnerships	
<b>Funding</b>	Identify and secure initial sources of funding	Start-up funding in place	Develop sustainable funding model
<b>Public Engagement</b>	Consultation and engagement	Embed Londoners in decision-making processes	Public involvement as a key strand of Demonstrator Projects
<b>Sponsor a portfolio of Health Improvement Demonstrator Projects</b>			
<b>Demonstrator Projects</b>	COVID-19 projects underway	Launch call for proposals	Projects underway, to deliver in 6-12 month timeframe
<b>Establishing a shared framework for a London Health Data Hub</b>			
<b>Discoverability of data assets</b>	Systematic approach to defining data mapping data flows	Agree meta-data standards. Work with HDR Innovation Gateway to onboard.	100 data assets visible through HDR Innovation Gateway
<b>Technical Infrastructure</b>	Detailed scoping – in partnership with NHS Digital and NHSX	Agree technical spec for interoperability Technology strategy and cloud strategy	Implement strategy
<b>Data curation</b>		Scope data wrangling requirements	Curation service in place
<b>IG processes</b>		Define and agree any shared review processes, in collaboration with national bodies	Publish clear rules and implement any new processes required.

## ANNEX B: A mandate from Londoners

The OneLondon partnership (representing local authorities, the Greater London Authority and the NHS in London) established a multi-stage process of public engagement, culminating in a deliberative event with 100 people over four days in February 2020.<sup>20</sup> The Citizens' Summit explored issues around trustworthy use of health and care data, to understand expectations and to explore trade-offs. It found that the public have a clear expectation that data should be consistently joined-up across London to support individual care, proactive care, planning and research.

The findings from the Citizens' Summit report have informed and guided this proposal. Key recommendations include:

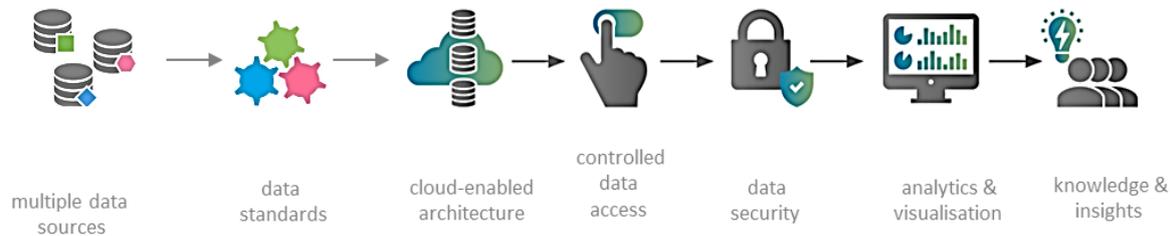
- Making better use of de-personalised data for planning health services was considered a “must-do” by this group of Londoners.
- De-personalised health and care data must be shared and used by relevant bodies to plan and improve services and demonstrably benefit health in London
- There should be trustworthy oversight of the system of joining-up and using health information, which should include ways for the public to be involved in ongoing policy development.
- Nine out of ten participants at the Citizens' Summit expected a single set of policies across all health and care organisations, with an expectation that data should be consistently joined-up.

We recognise that the Citizens' summit was completed before the pandemic, and so it will be important to continue the conversation in light of the shifting attitudes towards data use. Ongoing engagement will be essential to ensure a continuing mandate from the public, patients, healthcare professionals and institutions.

---

<sup>20</sup> <https://www.onelondon.online/wp-content/uploads/2020/07/Public-deliberation-in-the-use-of-health-and-care-data.pdf>

## ANNEX C: Framework for a London Health Data Hub



We propose to develop a single coordinated Health Data Hub, with nodes within partner Universities, serving the whole of London, the UK and beyond. The nodes will be created through formal collaboration agreements between NHS and Universities.

Design and delivery of the flexible research service will require significant multi-disciplinary specialist expertise, including a robust understanding of the source data and research uses. The ambition is that collectively the hub will have comprehensive coverage of most/all of London's population of 8 million individuals. There will be a clear remit to ensure coordination and drive improvements in health.

### 1. To make data assets visible

London's datasets will be made discoverable on the UK Health Innovation Gateway, ensuring the data is Findable, Accessible, Interoperable and Reusable (FAIR), through shared metadata and advanced discoverability tools. This will help demonstrate the richness of London's data assets to industry and investors. We envisage 100 datasets from up to 20 data providers could be available within a twelve month period.

### 2. Use Trusted Research Environments for analysis

Participants at the Citizens' Summit were clear that data must be held securely, with controlled access. In order to meet the 'safe setting' requirement, there needs to be a controlled environment where clinical, academic and commercial researchers can access data for research purposes. HDR-UK has published a Green Paper setting out the requirements for such a Trusted Research Environment (TRE).<sup>21</sup> This protects by design the privacy of individuals whose health data it holds, while still facilitating the types of large scale, high performance computing data analysis using that will bring insights into disease and improvements in healthcare. Individual level data would not be distributed in any form to researchers, but instead the TRE would provide access to a secure analytics environment (i.e. a safe setting) where researchers can bring analysis algorithms to the data.

A joint data architecture will be delivered to robust industrial standards, cloud-enabled and interoperable across the hub nodes. It will provide a secure, accessible data infrastructure with advanced data discoverability tools. Data controllers will be able to stay in control, while NHS, public and academic data will be accessible within the same environment. It will need flexible analytical workspaces, capable of ingesting and integrating diverse health data, adapted to different baseline capabilities, rapidly evolving data requirements and diverse user/industry needs. Annex D provides further detail about current TREs in London.

### 3. Provide expertise for Data Curation

There are also technical issues with lack of standardized data, and proprietary data formats which make interoperability difficult. Data must be accessible, interoperable, reliable and re-useable. High-quality data is essential and the Citizens' summit made a clear recommendation that that data being shared must be accurate.

<sup>21</sup> <https://ukhealthdata.org/projects/aligning-approach-to-trusted-research-environments/>

The Demonstrator Projects should be used to drive discussion about the implementation and use of standardised data sets. It will be important to improve knowledge and understanding of using consistent data terms, highlighting the benefits and opportunities it presents to referring clinicians, professionals, care providers, service users, and commissioners. The intention should be to embed the habitual use of these data standards, ensuring professional buy-in and institutional commitment across London.

The use of standardised information will improve data quality, ensure consistency and timeliness of data capture and extraction, allow integration of systems, and reduce the costs of interoperability. Dedicated resource for data curation and standardisation will therefore need to be a key element of the framework. There will also be opportunities for London to provide the expertise to contribute to, and collaborate with, the nation-wide approach to data curation currently being led by Office for Life Sciences and NHSX.

This will ensure we deliver interoperability and secure data provision across the nodes at scale, and are well positioned to contribute to a UK-wide and international network as part of the anticipated CSR investment.

#### 4. Ensure consistent rules about access

The Londoners in the Citizens' Summit concluded that, for data to be accessed and used responsibly, the 'five safes' framework should be used (see Box 1). Data must be shared for an agreed purpose, and only data relevant to the specific analysis must be shared. There should be legal penalties for misuse of the data and transparency, both about the decision making processes and approved uses of data, is essential.

##### Box 1: The 'Five Safes' framework\*

- **Safe data:** Is the risk of identification in the data minimised? Data must be depersonalised. Only data relevant to the specific analysis should be shared.
- **Safe projects:** Is the use of the data appropriate? Data must be shared for an agreed purpose, and only projects which serve public benefit should be allowed.
- **Safe people:** Can the researchers be trusted? There should be legal penalties for misuse of data.
- **Safe settings:** Is data accessed in an environment that is secure, accredited, monitored and limits unauthorised use? Trusted Research Environments should be used.
- **Safe outputs:** Are the results from analysis of the data anonymous when published?

\* For further detail, see blog by Office for National Statistics  
<https://blog.ons.gov.uk/2017/01/27/the-five-safes-dataprivacy-at-ons/>

To meet these requirements, and to provide clarity to data controllers, there must be clear rules about who can access data and for what purposes, and how decisions are made. At the moment, there are a number of existing local mechanisms for scrutiny of requests to access data. In some instances, joining up data may mean that access requests will need approval from several different bodies.

We will use the Demonstrators projects to test whether the interface between local mechanisms is working, or whether there need to be ways to streamline and simplify the approach. Does there need to be a London-level review panel to consider access requests? Individual data controllers could delegate authority to a central review panel, with delegated authority from Chief Executive

Officers and with public and patient involvement. However, local processes may be seen to be more trustworthy, and so it will be important to explore this in further detail with the public.

The Demonstrator projects will provide the opportunity to test the information governance, the review of safe projects and people, and the system-wide accountability, audit trails and checks that will be needed for a London data hub.

## **5. Embed public engagement**

The Citizens' summit recommendations emphasised the importance of the public having an ongoing role in policy development and be part of decision-making processes. As a first step, we recommend that the public and patients should be part of co-designing the Demonstrator projects and we will take this outline framework to patient groups, including the OneLondon panel and the HDR UK patient advisory group, to seek input and advice.

The Demonstrator Projects will be a key way to demonstrate how to embed public and patients in a meaningful way. We expect this to have two aspects:

- Clear communications and transparency will be fundamental to building a trustworthy approach.
- Involvement of public and patients in decision-making processes. In line with the Summit recommendations, we expect there should be a diverse citizen's advisory group to provide input on an ongoing basis and, in addition, two or three citizen representatives should be recruited to sit on decision-making boards with people who work in health and care as well as experts.

## **6. Establish fair partnerships**

NHS data is a valuable resource, and it is important to ensure that the NHS, and NHS patients, benefit directly from the innovations that are possible because of NHS data. It is therefore important that we reflect the need to consider fair partnerships from the outset.

Participants at the Citizens' Summit discussed the role of commercial partnerships, and there was general consensus that the NHS should be actively seeking to generate an income by realising the value of its health and care data through research and development. The expectation was for a fair and productive partnership, with benefits resulting from any research using patient data to be shared with the NHS, whether through shared Intellectual Property, royalties, a stake (in companies), profits and outputs (i.e. discount on new drugs). While Londoners concluded that London should be the first place to benefit from research using Londoners' data, they also wanted to ensure that the benefits could then be shared more widely across the NHS nationally.

These conclusions are very much in line with the guiding principles published by the Department for Health and Social Care to create a framework to help the NHS realise benefits for patients and the public where the NHS shares data with researchers.<sup>22</sup> The second principle requires that NHS organisations entering into arrangements involving their data, individually or as a consortium, should ensure they agree fair terms for their organisation and for the NHS as a whole. The third principle makes clear that any arrangements agreed by NHS organisations should not undermine or inhibit the ability of the NHS, at national level, to maximise the value or use of NHS data.

---

<sup>22</sup> <https://www.gov.uk/government/publications/creating-the-right-framework-to-realise-the-benefits-of-health-data/creating-the-right-framework-to-realise-the-benefits-for-patients-and-the-nhs-where-data-underpins-innovation>

The NHSX Centre for Improving Data Collaboration has been established to provide advice and guidance to NHS organisations when developing collaborations with industry or academia, and to help translate the principles into practice. The Demonstrator Projects will provide a useful test case to work with the Centre to think through the questions that need to be addressed, and to develop appropriate value sharing models, particularly where there are multiple data controllers. It will, of course, also be essential to ensure transparency about the purpose and nature of any partnerships.

London has the opportunity to embed best practice early, working closely with the NHSX Centre for Improving Data Collaboration. The nature of any commercial partnerships must be discussed openly from the outset to develop a framework for fair partnerships. The London Health Data Hub will provide the opportunity for academic/NHS/industry researchers will “breathe the same air” to research, co-create, test and adopt into the NHS (via public or commercial routes) data science innovation.

## **ANNEX D: Technical architecture**

### **Mapping existing Trusted Research Environments**

There are currently a number of infrastructure options to deliver a Trusted Research Environment (TRE) for London:

- ***OneLondon TRE capability: partnership with UK-SeRP***

Rather than establishing a local approach, the QMUL / Discovery team is testing the UK Secure eResearch Platform (UK-SeRP), developed by Swansea University, as a secure virtual analytic platform. This decision was taken to avoid duplicating effort and resource, and because the team does not have the developed expertise. UK-SeRP is a software offering that provides a secure, collaborative environment, allowing research groups to access complex linked datasets while meeting best practice in data management, security and information governance. Data owners and researchers can store, access, share, analyse and link data at scale in a governed environment while maintaining full control of the data. The partnership will initially develop a TRE for North East London Discovery Programme approved research projects, with the aim in the longer term for Discovery Collaborative projects covering two or more STPs participating in OneLondon. They will also develop and test governance and data access approval processes and systems.

***Imperial College Health Partners: Discover NOW / HDR Hub for Real World Evidence***

The NWL Whole Systems Integrated Care (WSIC) database, created by the Collaboration of NWL Clinical Commissioning Groups, provides a single repository of de-identified longitudinal data which links primary, secondary, mental health and social care data for over 2 million people and 362 GP practices in NWL. Discover, the health research platform, has been developed by ICHP and the NWL CCG Collaboration provides researchers with a single point of access to data and analytics tools.

<http://integration.healthiorthwestlondon.nhs.uk/>

Other examples of TREs in London include:

- The Clinical Record Interactive Search system (CRIS) at the NIHR Maudsley Biomedical Research Centre (BRC) KCL - CRIS Maudsley EHR - TRE - login and run NLP analytics with good oversight mechanisms
- Great Ormond Street Hospital NHS Foundation Trust (GOSH) research and innovation platform, provided by Aridhia
- The Biomedical Research Centres (BRC) are also introducing TREs through membership of the UK Health Data Research Alliance.

We do not recommend prescribing a single commercial or academic TRE solution for London at this stage. Instead, it would be more appropriate to use the Accelerator Projects to move towards shared standards and to build interoperability across distributed platforms through the use of open APIs. It will be particularly important to explore the possibility of federated access between different TREs, rather than moving data around the system.

Using a flexible approach will ensure the use of data is agile, timely and cost-effective, while supporting innovation in the system. By ensuring consistent rules and standards, it will be possible to develop greater interoperability and potential rationalisation over time, and to move towards a sustainable system.

## Learning from other industries

Other industries have digitally instrumented **data ecosystems** that has led to a new way of thinking about cross-organisational working, productivity and impact, even when Institutions have simultaneous collaborative and competitive relationships<sup>23</sup>.

In the past, physical assets and internal processes created barriers to entry. Today, data and technology capabilities are part of the new competitive “intangible assets” while traditional resources can easily be bought and sold in market ecosystems. In the past, the fact that a firm owned its own computers and ran its systems in a unique way gave it an advantage.

Computing capacity can be acquired on a subscription basis from AWS, Microsoft’s Azure, Google’s Cloud and numerous other platforms. These compute platforms are also delivering scalable research and innovation data services to offer a high quality user experience.

APIs (application programming interfaces) and advanced data capabilities provide an opportunity to establish new and more tightly integrated, seamless working relationships between organizations, meaning that each player can perform those functions they do best in cooperation with their data ecosystem partners. Digital and data-driven ecosystem relationships are complex and require new architectures, governance models, commercial and operational contracts. This has however been achieved in other industries with Open or API banking (eg SWIFT - vast messaging network used by banks and other financial institutions to quickly, accurately, and securely send and receive information, such as money transfer instructions.) and telecommunications (GMS).

---

<sup>23</sup> <https://www.i-cio.com/management/insight/item/the-swift-rise-and-surprising-importance-of-digital-ecosystems>

## ANNEX E: Examples of potential Health Improvement Demonstrator Projects

Theme	Accelerator Project	Partners	Questions to be Addressed
Cancer	Impact of Cancer Diagnosis & Interim Treatment Measures	<ul style="list-style-type: none"> <li>* AHSN Network</li> <li>* Data-CAN</li> <li>* London Cancer Alliance</li> <li>* NIHR ARC North Thames &amp; Cancer Policy Research Unit</li> <li>* UCL Institute of Epidemiology &amp; Department of Applied Health Research</li> </ul>	<ul style="list-style-type: none"> <li>* Determine the impact of COVID-19 on referral and diagnosis of bowel cancer patients – studied with datasets that are not readily available in real time at a national level but can be accessed directly from providers and collaborators across London to provide timely actionable intelligence.</li> <li>* Evaluate the impact of uptake and use of the permitted interim COVID-19 treatment guidelines for cancer patients across a number of cancers (can be extended to the national (England) dataset).</li> </ul>
Mental Health	Recognising and monitoring mental health sequelae of Covid-19: Identification, surveillance and care planning	<ul style="list-style-type: none"> <li>* Discover-NOW (including partners Imperial College, NWL CCGs and Local Authorities)</li> <li>* Future: Other STPs as data becomes available in OneLondon</li> </ul>	<ul style="list-style-type: none"> <li>* Does Covid-19 infection lead to increased physical morbidity and mortality in patients suffering from mental illness?</li> <li>* Is this related to pre-existing cardiovascular, respiratory or metabolic diseases?</li> <li>* What are the prognostic outcomes of patients who survive Covid-19 infection after an initial sic month period in terms of continued physical and mental health sequelae? Does smoking confer a protective factor in patients with SMI?</li> </ul>
Childhood Obesity	Proposal In Development		
Vaccinations	Digital Innovation for Equitable Delivery of Timely Childhood Immunisations	<ul style="list-style-type: none"> <li>* NHSE/I</li> <li>* STP / ICS Partners</li> <li>* LSHTM</li> <li>* UCL Institute of Child Health</li> <li>* Public Involvement</li> </ul>	<p>Does the Active Patient Link Immunisation (APL-Imms) tool improve timely and equitable delivery of childhood immunisations in general practice in London? Specifically:</p> <ul style="list-style-type: none"> <li>• Does it increase the proportion of children completing primary immunisations by age 6 months?</li> <li>• Does it increase MMR timeliness as measured by proportion receiving MMR by age 15 months?</li> <li>• Does it reduce unwarranted variation in timeliness by CCG, deprivation and ethnicity?</li> </ul>
	Childhood immunization coverage, ethnicity and deprivation in Hammersmith and Fulham	Currently working with a new H&F Immunisations working group reporting to the HWBB chaired by the DPH, but trying to reconcile data sources	<ul style="list-style-type: none"> <li>* Does analysis of child immunisation data show a relationship between ethnicity and immunisation coverage, or is this explained by deprivation?</li> <li>* What evidence is there of a relationship with socioeconomic status, and other plausible mediators?</li> </ul>
Public Health Priorities	An Exemplar for CVD in London	<ul style="list-style-type: none"> <li>* AHSNs</li> <li>* BHF</li> <li>* AHSCs</li> <li>* Applied Research Collaboratives</li> </ul>	Support for Quality improvement in primary care management of the high-risk conditions for CVD through the development of search tools, prompts and dashboards at a PCN and Practice level utilising the rich data from CVDPREVENT
	Dementia Early Diagnosis & Prevention	<ul style="list-style-type: none"> <li>* UCLP</li> <li>* ICHP</li> <li>* QMUL</li> <li>* ELFT</li> <li>* Barts Health</li> </ul>	<ul style="list-style-type: none"> <li>* What are the determinants of brain health and dementia risk most relevant to the heterogenous populations of London?</li> <li>* How can these be used to deliver targeted interventions to prevent and improve early diagnosis in individuals at high risk?</li> </ul>

Theme	Accelerator Project	Partners	Questions to be Addressed
COVID-19 Impact on Patients	Impact of Covid-19 on preterm birth rates and outcomes: Informing local and national policy	<ul style="list-style-type: none"> <li>* Discover-NOW including partners Imperial College (Neonatal Data Analysis Unit, Small Area Health Statistics Unit)</li> <li>* NWL CCGs and Local Authorities</li> <li>* Future: Other STPs as data becomes available in OneLondon</li> </ul>	<ul style="list-style-type: none"> <li>* Does analysis of large, population-based datasets confirm a fall in preterm births during lockdown?</li> <li>* What evidence is there of a relationship between preterm births and outcomes with air quality, sociodemographic and socioeconomic status, ethnicity and other plausible mediators?</li> </ul>
	Surveillance of Covid-19 in London and identification of unexpected long-term health complications	<ul style="list-style-type: none"> <li>* Discover-NOW partners including Imperial College &amp; NWL CCGs</li> <li>* OneLondon</li> </ul>	<ul style="list-style-type: none"> <li>* How can COVID-19 cases be best identified data, and what patterns (including in demographics, service use and outcomes) are evident over time and geography?</li> <li>* What factors impact chances of COVID-19 infection, and outcomes; including and especially ethnicity, age, gender and existing long-term conditions?</li> </ul>
	Ethnicity, inequalities and Covid-19 in Hammersmith and Fulham	<ul style="list-style-type: none"> <li>* London Borough of H&amp;F</li> <li>* Imperial College</li> <li>* WSICC</li> <li>* PHE</li> <li>* Future: Other STPs as data becomes available in OneLondon</li> </ul>	<p>Does analysis of large, linked population-based datasets find any relationship between overcrowded housing, deprivation, ethnicity and risk of Covid-19 and also death from Covid-1?</p>
Other	Coordinate My Care: Bringing together data to evaluate the impact of digital Urgent Care Plans	<ul style="list-style-type: none"> <li>* CCGs (SWL Lead)</li> <li>* EoL Clinical Network</li> <li>* Dementia Clinical Network</li> <li>* NHSE / I</li> <li>* AHSN Network</li> <li>* Coordinate My Care</li> </ul>	<ul style="list-style-type: none"> <li>* What is the overall cost of care for patients who have an electronic Urgent Care Plan (UCP) vs patients who don't have a electronic UCP.</li> <li>* What is the difference in the type and range of services (Acute, Community, hospice etc.) who support patients who have a digital IUCP vs patients who don't.</li> <li>* Do patients who have a digital IUCP have better outcomes vs patients who don't have an IUCP?</li> <li>* What are the characteristics (age, condition, ethnicity) of the patients who most benefit from having an IUCP?</li> </ul>
	London Informatics Consult Accelerator: Demonstrating New Capability in Learning Health Systems	<ul style="list-style-type: none"> <li>* Patients</li> <li>* Clinicians</li> <li>* NHS Trusts</li> <li>* Health Data Alliances partners</li> <li>* Digital Innovation Hubs</li> <li>* International partners</li> </ul>	<p>The Accelerator seeks to establish in London the capability to return new evidence to make better treatment decisions. Not after months of academic deliberation, but clinically reviewed, quality assured evidence in clinical timescales (e.g. 48 hours)–this is the central idea of the 'Informatics Consul</p>

## **ANNEX F: International comparators**

### **Mayo Clinic: Clinical Data Analytics Platform**

The Clinical Data Analytics Platform (CDAP) is a strategic initiative of the Mayo Clinic to improve health care through insights and knowledge derived from clinical data, by applying machine learning approaches to de-identified data. The timetable for the project has been accelerated significantly in response to the COVID-19 pandemic.

The CDAP model is based on three pillars:

- All data are de-identified and aggregated to ensure the protection of patient privacy.
- AI/ML tools and analytics are brought to the data. All analysis is conducted in a safe Mayo Clinic Cloud (MCC) environment so that no data leaves Mayo, even when external partners use the data
- The MCC environment is monitored and audited to assure no misuse of the de-identified, aggregated data.

In September 2020, Mayo announced completion of the deidentification of structured portions of 10 million Mayo Clinic patient records, along with 2.5 million unstructured portions. This de-identified data is already available through CDAP.

Insights derived from the deidentified data will be available to Mayo Clinic research staff and as a service to the biopharmaceutical industry to identify targets for new drugs, find potential new uses for existing drugs and optimize clinical trials. The intention is that the combination of curated, deidentified clinical data and advanced analytics technology will accelerate scientific discovery and development of new therapeutics.

### **Nordic Commons**

In 2019, NordForsk published a report setting out proposals for a Nordic Health Data Commons that allows secondary use of sensitive health data in research, health care and industry. The Nordic Council of Ministers have now agreed funding from 2021 to develop the proposals.

The aim is to combine data assets, including longitudinal, registry and biobank data, from Denmark, Finland, Norway and Iceland, with a combined population of 27 million people. The starting point for the report was concerns that the data was underused, and that the value would be lost unless data was made more easily accessible with increased cooperation.

The concept of a Nordic Commons is based on existing national initiatives. The key focus is to develop technical solutions for secure cross-border transfer, access, and analysis, supported by a coherent legal and ethical framework. Data must be FAIR, and there is an emphasis on enhancing the quality of data.

The report has four main pillars:

1. A Nordic federated secure platform for processing sensitive personal data – a Nordic Health Cloud.

2. Nordic health data described with rich metadata according to the FAIR principles – a Nordic Health Metadata Repository .
3. A coherent legal and ethical framework.
4. A research funding programme for technology and competence development.

The goal is to establish a Nordic Commons in the form of a federated, secure, scalable environment for using Nordic sensitive health data sets in research. The report recognises that implementation will need to be incremental, due to the diversity of stakeholders and data. The intention is to use stepwise implementation, adapting the technology to different scenarios and use cases. Competitive calls for research funding will be issued that explicitly target technology and infrastructure development.