

HDRUK
Health Data Research UK

UK Health Data
Research Alliance

Health Data Research
Innovation Gateway

REPORT:

All of us together

UK Health Data Research
Alliance Symposium

Tuesday 1 December 2020

All of us together: UK Health Data Research Alliance Symposium

1 December 2020

Executive Summary

COVID-19 has brought the UK health data science community together as never before. It has been extraordinary to see the what has been achieved by so many partners working together: more than 530 datasets in the Health Data Research Innovation Gateway searched more than 10,000 times in the past month; Health Data Research Hubs fulfilling more than 200 contracts within industry; multiple large COVID-19 datasets linked and made safely accessible in Trusted Research Environments; and thousands of patients and members of the public engaged and involved in shaping these approaches in trustworthy ways.

The pandemic has highlighted the role that health data can play in understanding this new disease and saving lives, from genomics and epidemiology through to clinical trials of treatments. But there is much more to life - and health - than one virus.

This year has also accelerated changes and conversations that were already starting to take place, including ever greater reliance on digital technologies and connectivity, and a sharper focus on ensuring that everyone in the population is represented in health data research and can benefit from it.

The virtual format meant that more than 500 people signed up to join the event from across the health data science community, including the NHS, industry, universities, charities, government, patients and the public, with speakers from as far afield as the west coast of the US and Thailand.

The [packed one-day agenda](#) explored the vast and exciting possibilities of using health data at scale to improve people's lives and healthcare, here in the UK and internationally, and examined what's needed to expand our knowledge and tools to be fit for the future.

Several key themes emerged over the day, all of which are active areas of work for the Alliance and discussed in more detail below:

- **Collaboration:** The importance of working together as a diverse community to create value from health data that benefits all
- **Security and access:** How technologies such as Trusted Research Environments and federated approaches are enabling secure access to health datasets
- **Data diversity and standards:** The importance of gathering diverse datasets that are representative of the population, and working to shared open data standards

- Patients and the public: Building and maintaining public trust and consent for health data research through transparent communication, engagement and consultation

There's an old proverb that says, *"If you want to go fast, go alone; but if you want to go far, go together."* This year has shown that we can move fast and go far by working together with urgency to address the greatest public health threat in a century. We've seen research projects set up in a matter of weeks that would take months or years at any other time, without sacrificing security or public consultation.

Let us take this energy and lessons learned forward into the post-pandemic world, working together to unite and improve health data at a scale to respond to global health challenges – both today and in the future. Together, we are making history.

"The pandemic has highlighted and amplified the issues that were already there and offers an opportunity to reset the system going forward."

Ottoline Leyser, Chief Executive, UK Research & Innovation

"There are no better career opportunities today than in healthcare data science, solving real-life problems... Health data science can take the lead in transforming society."

Martin Levermore, Chair, External Advisory Board, PIONEER - The Health Data Research Hub for Acute Care

>>

>>

"This year has forced radical behaviour change. There is an opportunity for us to look at lessons learned and decide whether there are some changes we'd like to take forward."

Douglas Gurr, Director, Natural History Museum and Former Managing Director, Amazon UK

>>

"If the last decade was all about digitising healthcare, the next decade is all about making that data good for something, through the power of computing and machine learning."

Laura Robinson, Health & Life Sciences Lead, Microsoft

Collaboration is key

While the 2019 Alliance Symposium focused on bringing together the UK health data community, this year's meeting not only showcased the work of Alliance partners within the four nations but looked across sectors and international borders to gain new insights.

We saw some impressive examples of collaboration, both here in the UK and with international partners. One notable example is the PRINCIPLE trial testing treatments for older people in the community. By bringing together a multi-disciplinary team including governance and technical experts including from NHS DigiTrials, Patient Engagement Experts and the Information Commissioner's Office, the trial team were able

to create an approach that identifies eligible participants for trials within 24 hours of receiving a positive test result in the community.

As well as growing and strengthening the existing connections within the rapidly growing field of health data science, we urgently need to bring more people in, including ‘research ready’ scientists and data managers. This requires a diverse workforce with a wide range of skills, talents and perspectives drawn from multiple sectors working together to realise the value and benefits of health data science. To support this aim, we have launched a [partnership with #10000 Black Interns](#), offering paid internships to young Black data science students across 20 organisations within the health data community.

“HDR UK has a major role to play in enabling collaborations... If you bring researchers together to look at data, they will work together”

Cathie Sudlow, Director, BHF Data Science Centre

Security and access

Secure access to data was another recurring message. This is important work all driven by many Alliance members. COVID-19 has accelerated work that was already ongoing: around making data more accessible, about streamlining processes, about facilitating access for researchers. Custodians have demonstrated it is possible to provide access to health data efficiently and have enabled fast generation of insights which has been crucial in tackling a public health emergency.

Where appropriate, data should be made openly accessible. And even if data can’t be freely shared, then open-source tools and methods can. [Trusted Research Environments](#) – secure computing environments that allow researchers to come to the data, rather than the other way round – are emerging as an important solution to the challenge of providing secure access to health datasets, with a number of national and local TREs now available across the UK.

Several sessions discussed the concept of federated research, allowing researchers to run analysis on datasets held in multiple secure environments at once. A similar federated approach can be applied to researchers themselves, offering secure, validated credentials similar to a ‘passport’ and ‘visas’ in order to access data. For more information, see [this video from the Global Alliance For Genomic Health](#), shared by Ewan Birney, Director, European Bioinformatics Institute, as part of his talk on standards and access to data for genomic medicine.

We also heard about exciting progress on the [HDR UK’s Innovation Gateway](#), an online portal for the discovery of the UK’s health datasets, as well as a collection of publications and tools for data analysis. This progress has only been possible thanks to the efforts of Alliance members to improve the quality, access and use of datasets that are available through the Gateway.

Researchers are now able to apply directly through the Gateway to securely access data from eight major custodians, moving this resource closer to delivering on its promise of being the principal destination to access health data research. There are currently over 530 datasets accessible from the Innovation Gateway, with this figure continuing to grow rapidly.

Data access procedures are speeding up, but there is always room for improvement without sacrificing security. One solution for streamlining the process is increased automation, although this relies on high-quality datasets with good metadata. A number of speakers also highlighted the need for long-term investment in data science infrastructure to build a sustainable health data research ecosystem, together with support for often-overlooked areas such as governance and data standards.

“We need to think what’s practical in the short-term while being mindful of the broader ambition.”

Dave Robertson, Interim Director, HDR UK Scotland

Data Diversity and Standards

Diversity of data was raised as an issue across many of the Symposium sessions: if the outputs of health data research are to be relevant to all, they must be developed using datasets that are representative of the population. This is not only true of the data that is used to develop algorithms and data-driven solutions, but also of the data that is used for training and validation.

The Alliance is [committed to collecting diverse and inclusive datasets](#) to address the questions that will come next. For example, more could be done to uplift routine operational health data to research-grade data, ideally aiming to generate ‘research ready’ data as part of routine healthcare.

Increasing attention is being paid to gathering ‘real world’ health data, capturing information about people throughout their lives rather than just interactions with the health service at the point of need. This includes information about quality of life, lifestyle and living conditions, user-generated data from apps and wearables such as fitness trackers, and data from the UK’s numerous rich longitudinal cohort datasets. This data is complex and hard to extract, and often voluminous, but being able to gather and analyse it will shed much light on both sickness and health.

>>

“It’s about improving the entire end to end journey for data, about how we collect it, process it, curate it, link it, establishing the data quality and then finally making it available for research – doing that, we can create a research data ecosystem that can be sustained beyond the existing pandemic.”

Garry Coleman, Associate Director of Data Access, NHS Digital

>>

Linking datasets together securely from multiple sources and locations remains a challenge, particularly when including data from non-healthcare settings such as local councils. We need a consistent, informed approach to data linkage that goes beyond using NHS numbers, along with access to information governance experts who can advise on appropriate and ethical channels to enable data flows. Agreed international data standards and access will also become increasingly important as global collaboration in health data research increases. One example where this is happening is the [International COVID-19 Data Alliance \(ICODA\)](#), convened by Health Data Research UK.

Data standards, quality and interoperability are also vital for machine learning and AI approaches, which offer great potential for scaling and accelerating the pace of health data research, particularly in combination with cloud computing. However, these ‘data hungry’ applications rely on high quality, large, representative datasets for development, training and validation, so attention must be paid to collecting and labelling appropriate datasets and making them discoverable and securely available for this kind of work.

“Metadata is data, and it needs the same kind of care as you give to the actual values of the data itself.”

Mandy Chessell, Distinguished Engineer, IBM

Patients and the public

Public and patients are enablers and important advocates of health data research, and informed consent and transparent communication must be at the heart of this work. Public trust in health data science is easy to lose and hard to regain, but it is essential if we are to move forward. As a community, we should always be thinking how to communicate the benefits of health data research and explain how it works while avoiding hype.

The importance of investing time and effort in patient and public consultation was raised many times over the course of the day. This is essential not just here in the UK but globally, particularly in low and middle-income countries, to make sure that health data research works for all. There is a need for a conversation about agreed international standards for patient and public engagement and consultation, similar to the conversation around international data standards, which requires investment and support from funding organisations.

Furthermore, data technologies and tools should be built with the patient in mind, starting with consent, privacy, utility and patient benefit. It is notable that all four Lightning Talks were heavily focused on patient benefit. Liz Dobson, CEO of the IBD Registry won the session in recognition of her work building an COVID-19 risk tool in just 8 days. This [simple online tool](#) empowered more than 34,000 IBD patients to understand their own risks and communicate it back to their healthcare provider – a shining example of a patient-centric approach with real world usefulness.

“Maintaining trust of the patients and public should be paramount. People are understanding when humans make mistakes. People are also understanding when computers make mistakes, but we need transparency to explain why.”

Cian Hughes, Informatics Lead, Google Health Research

“We should always look at data through the lens of the patient... listen to what they are telling you.”

Amitava Banerjee, Associate Professor in Clinical Data Science and Honorary Consultant Cardiologist, University College London Hospitals NHS Trust

About the Symposium

Convened on behalf of the [UK Health Data Research Alliance](#), the second annual Symposium was an opportunity to celebrate the transformation of health data research infrastructure in the UK over the past year and reflect upon the impact this is already having on our knowledge, understanding and response to COVID-19 and other conditions.

The Health Data Research Alliance is a partnership of more than 30 of the UK’s leading health, care and research organisations working together to establish best practice around the ethical use of UK health data for research and innovation at scale.

Many representatives of Alliance members actively participated in the Symposium as speakers, chairs, part of the advisory group or in supporting roles and we are grateful to them all. It was impressive to see the evolution that has happened since last year’s Symposium, thanks to the dedicated work of numerous partners who have put in practice many of the ideas developed within the Alliance.