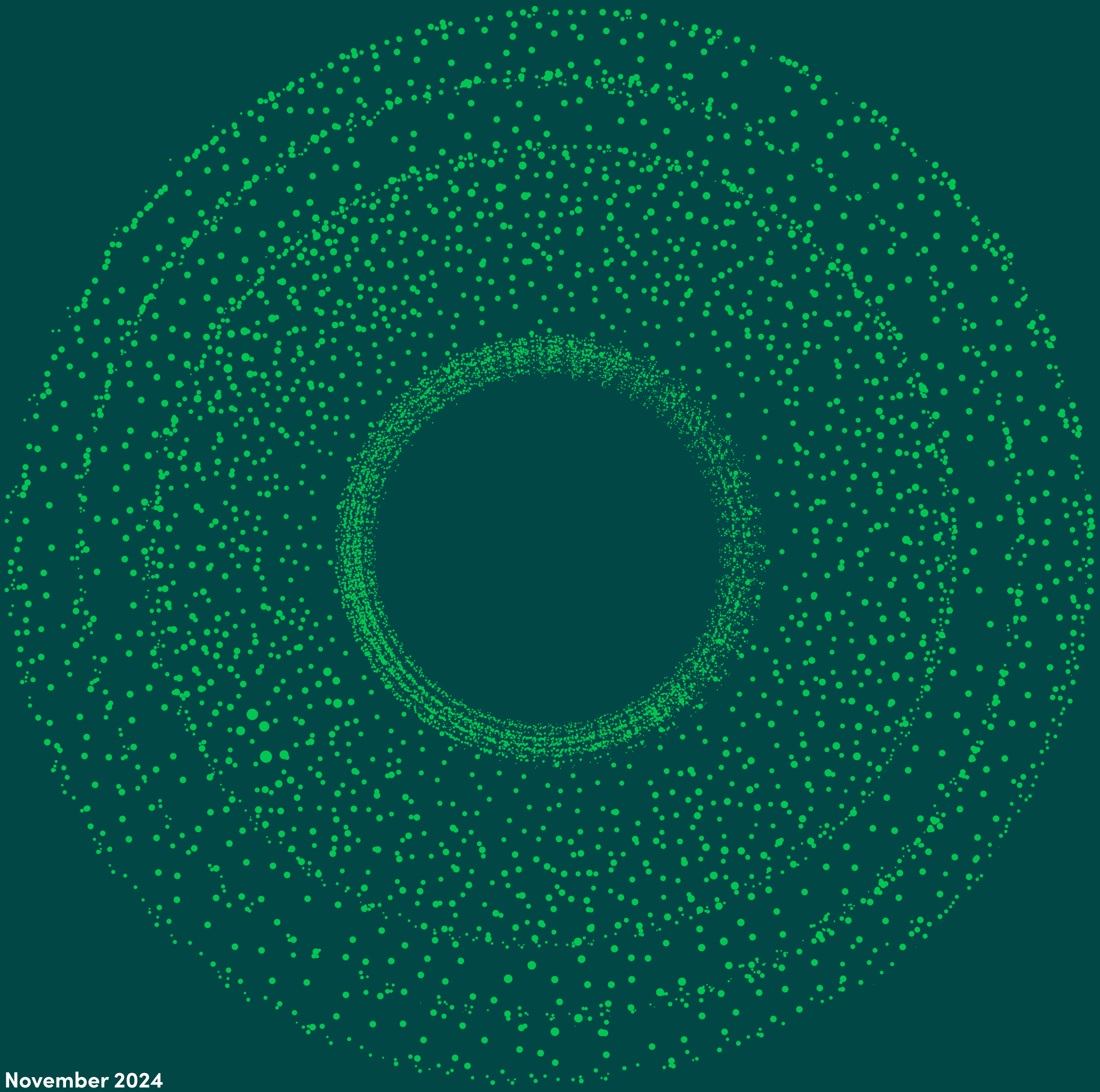
Uniting the UK's Health Data: A Huge Opportunity for Society

A review of the UK health data landscape commissioned by the Chief Medical Officer for England, the UK National Statistician and NHS England's National Director for Transformation

Executive Summary



November 2024

Personal note from the author

Undertaking this review has been both a daunting task and a great privilege. It has been hugely rewarding to discuss the UK-wide health data landscape with hundreds of people who have generously provided their time, expertise, knowledge, experience, views and advice. I am very grateful to each of them.

Through my various roles, I have been embedded in the UK's health data landscape for many years and I thought I knew it well. But conducting this review has reinforced just how extraordinarily complex this landscape is – one where finding potential solutions to the toughest challenges requires a broad understanding not only of data, science and technology, but also of the health and care system, of government, and of ethical, legal, social, cultural, behavioural, financial, geographical and political factors. Further, the sheer volume of information about the wide range of health-relevant data sources can sometimes seem like a bottomless pit. This is compounded by variation between the UK's four nations as well as by frequently changing organisational labels, structures, strategies and policies. Readers with a deep knowledge of any part of the landscape will inevitably find gaps in this review and for those I apologise.

In my roles as a doctor, and as both a user and creator of large-scale health data resources for research, I have seen the huge benefits for patients and society from a wide range of uses of health-relevant data. But I have also experienced, repeatedly over many years, the frustration of knowing that there are vast amounts of such data that could and should be – but are not – accessed and used to improve patient care, and to advance health research, care and policy for patient and public benefit. Consulting widely in preparing to write this review has confirmed that I am not alone: this frustration is widespread across the NHS, our academic institutes and universities, charities, organisations representing patients and the wider public, the life sciences industry, research funders and government.

Across the UK, despite key advances in recent years, we are simply not maximising the benefits to society from the many already-existing sources of health-relevant data. In some respects, we are even slipping backwards from some of the excellent progress made during the COVID-19 pandemic in broadening safe, secure access to and use of health-relevant data for patient and public benefit. Far too many lives are unnecessarily lost or ruined because of blockers or delays in safe and secure access to, linkage and analysis of existing health data. These blockers impede the generation of insights to guide and improve our health and care system. They delay or prevent hundreds of medical and population health research studies that collectively involve millions of people across the UK. These studies are essential to improving our health and wellbeing, through their aims to unravel the underlying causes of diseases; to develop better diagnosis, prevention and treatment strategies; to test these in clinical trials; and to undertake analyses in whole populations to assess their uptake, effectiveness and safety in the "real world."

There is a huge opportunity to turn this backward slippage around, to capitalise on the UK's substantial health data assets, and to accelerate towards a future where the power of health-relevant data is fully realised. This will only happen if we work collectively across the UK to simplify the overly complex landscape and to lay out a coherent vision and roadmap, where benefit for patients and the wider public is the key motivating goal.

Cathie Sudlow

Professor Cathie Sudlow, OBE

Key points

- Every day, health and care professionals, researchers and policymakers use health data safely to improve people's health and lives.
- People in the UK overwhelmingly support the use of their health data, with appropriate safeguards, to benefit themselves and others.
- The UK has abundant sources of data relevant to our health, both from its unique National Health Service and a range of other sources.
- The most powerful insights come from linking the different sources of data together.
- But health and care professionals, researchers and policymakers face many obstacles and delays in accessing, linking and analysing health data to improve people's care and lives.
- These barriers arise from the UK's complex and inefficient systems for managing and accessing health data.
- They prevent health and care professionals from accessing all the information they need to provide the best patient care.
- And they prevent or delay crucial analysis and research about health conditions affecting millions of people across the UK.
- We are letting patients and their families down as a result.
- We need to recognise our national health data for what they are: critical national infrastructure that can underpin the health of the nation. They should be treated as such with a strategy, leadership and investment to match.
- This review sets out how that can be achieved with five key recommendations.

Executive summary

All aspects of our lives are increasingly captured in digital form, and the health and care system is no exception.

An abundance of data is generated each second of every day through our many encounters with the National Health Service (NHS) when we visit our general practitioner (GP), spend time in hospital or have a blood test or scan. There are also data relevant to our health from social care, education, justice, earnings and disability settings, not forgetting data from those who take part in population or clinical research studies and from constant monitoring of the weather and pollution levels, all while our phones and devices measure step count, heart rate and sleep patterns.

We have a huge opportunity, and indeed a responsibility, to use health data safely and securely to improve health, wellbeing and prosperity across society.

Health data can be accessed for uses beyond our own direct care, for example when people give informed consent for use of their data, or when approved studies for public benefit access large datasets that have been stripped of patient-identifying information and held in protected environments.

Lord Darzi's recent review of the state of the NHS in England[1] highlights the critical condition of our health and care system. He calls for a major 'tilt towards technology' as one of the ways to improve the NHS, and points to the potential of artifical intelligence and life sciences breakthroughs to transform care and treatments. These advances will rely on the effective and trustworthy use of health data.

As **Chapter 1** of this review sets out, health data can and should be used to:

- support the care of each one of us if we become sick;
- inform an intelligent health and care system capable of predicting and responding to varying demands, for example managing outpatient and operating theatre waiting lists and addressing bottlenecks in hospital emergency departments and in discharges from hospital;
- support the planning and equitable delivery of health, care and public health services that meet the needs of local, regional and national populations, keeping people healthy through preventing ill health as well as looking after them when they are unwell;
- enable a wide range of research and innovation to benefit patients and the public.

Improving and saving lives

The four nations of the UK have a long history of collecting national databases of health-relevant data from their entire populations, going back as far as the 1950s in Scotland. Almost all the 67 million people living in the UK receive most of their healthcare from the NHS. This makes our national collections of data among the largest and most comprehensive worldwide.

As a result, the UK hosts some of the best examples globally of transformational use of health data for public benefit.

- UK Biobank is a large-scale database and research resource used by over 30,000 researchers worldwide to better understand the causes and consequences of many different health conditions, such as heart disease, stroke, cancer, diabetes, arthritis, mental health conditions, dementia and many others. It is also used to develop new approaches for their prevention, treatment and diagnosis. More than 500,000 volunteers have undertaken extensive questionnaires, measures and imaging, donated samples and given permission for their health to be followed through their routinely collected NHS and other health-related records. This means UK Biobank has unparalleled depth and breadth of data and samples for carrying out high-quality research.
- Building on the success of UK Biobank, Our Future Health aims to be the UK's largest ever health research programme, with a target of five million volunteers. The programme is designed to support multiple research initiatives to discover and test more effective approaches to prevent, detect and treat diseases. It has partnered with national NHS organisations such as NHS England, using centralised NHS databases to issue invitations to take part to millions of eligible people across the UK. As a result, well over one million volunteers have already joined the programme. Their health and wellbeing will be followed over many years through linking to national health-related records from the NHS and many other sources.
- During the pandemic, UK policymakers were rapidly informed about the impact of COVID-19 infection and vaccination on people with different health conditions and of different ages, ethnicities and socio-economic circumstances. This was made possible through the secure linkage and analysis of a variety of health datasets for the whole populations of the four UK nations.

However, these major UK-led health data-driven successes are far too often the exception rather than the rule. The complexity of our data systems and data governance means such work is far from routine when it could be business as usual.

The national response to the COVID-19 pandemic drove remarkable progress in broadening secure access to health-relevant data for patient and public benefit. The data-driven RECOVERY trial is a fantastic example of this. It was able to answer key questions about how to treat severe COVID-19 and led to the widespread use of treatments such as dexamethasone, saving hundreds of thousands of lives worldwide. Yet some of these advances in the use of data to inform healthcare are now falling back to pre-pandemic approaches.

Existing national data collections should be relatively straightforward for approved analysts to access, link and analyse. But in practice, access is difficult, slow or impossible. For example, access to data currently available via NHS England often takes many months or even years.

---

[1] See https://www.gov.uk/government/publications/independent-investigation-of-the-nhs-in-england

Furthermore, many national NHS England data are only accessible for COVID-related analysis and research (for example general practice data and national cardiovascular audits) but not to tackle other health conditions, such as other infectious diseases, cancer, heart disease, stroke, diabetes and dementia.

And more complex types of health data generally do not have national data systems (for example most laboratory testing data and radiology imaging). Existing examples of what can be done (for example Scotland's national medical imaging database) make national solutions for these complex data ambitious but achievable goals.

The existing barriers impede work to guide and improve our health and care system. They delay or prevent hundreds of health research studies that could:

- improve our health and wellbeing by unravelling the causes of diseases;
- develop better diagnosis, prevention and treatment strategies for conditions affecting many millions of people;
- test these in clinical trials and assess their uptake, effectiveness and safety in the 'real world'.

Support from patients, public and professionals

People in the UK overwhelmingly support the use of their health data to benefit themselves and others. Surveys, in-depth focus groups and other information-gathering exercises over the last 15–20 years have consistently shown this, as summarised in **Chapter 2**.

Most people want to know how their data are being and will be used. They are concerned to know how their privacy will be protected and their data kept secure; that robust and transparent mechanisms are in place to ensure that data are used for public good; and that they can choose not to have their data used for certain purposes beyond their direct clinical care.

Many people want to be able to access and, where needed, amend their own health records. And many are more cautious about uses of data from which organisations might profit financially.

Less is known about the views of healthcare professionals. General practice data are some of the most important for improving healthcare. But, while some GPs are at the forefront of efforts to ensure data are used widely for patient and public benefit, we know that others are more circumspect than patients and members of the public. Some GPs have concerns about inadvertently breaching laws that protect the confidentiality and privacy of patient data. Some may also be concerned about data being used for performance management. However, despite these concerns, GPs are generally supportive of wider uses of health data for patient and public benefit. It is important that measures to increase the use of data reduce, or do not add to, the burdens of general practice.

The complex health data ecosystem: data from many sources, not just the health service

Fulfilling the potential of health data to improve lives is not straightforward. It can be beset by delays within a system that is frustrating to navigate. This is caused by the complexity and fragmentation of the health data ecosystem.

To start with, the NHS is not one organisation but many – general practices, hospital and mental health trusts, integrated care boards, and more – and its constituent parts do not always work together effectively.

The digitisation of healthcare in the UK lags behind other high-income countries and other sectors. Indeed, parts of the health and care system, social care in particular, still depend on paper records. The piecemeal introduction of many different computer systems into the health and care system, provided by many different companies, has created difficulties in 'interoperability' – the ability of these different systems to talk to each other. People working in multiple NHS bodies spend a lot of time maintaining the many contracts needed between their organisations and multiple commercial computer system suppliers.

Added complexity comes from the statutory and common law frameworks that govern access to, and uses of, health-relevant data. These are complicated to start with and are interpreted and applied differently by the many data custodian organisations across the complex ecosystem. Furthermore, the common law position and mechanisms for complying with it differ between the four nations of the UK.

Amid this complexity, this review was commissioned to map the health data landscape across the UK. **Chapter 3** is by far the longest in the review and provides a guide to the many sources of health-relevant data: the health data collected and held by general practices, hospitals, laboratories, X-ray and scanning departments, high street opticians, pharmacies, dentists and others, as well as the health-relevant data coming from many sources beyond the NHS. We hope this chapter will prove useful to all those using data to improve healthcare, whatever their perspectives on how it should be accessed.

Each of the various sources of health-relevant data can separately provide useful information. But their power comes from when datasets are linked together. This is when the most important and transformational insights emerge, as **Chapter 4** demonstrates.

For example, we can only really know if breast cancer screening is improving cancer outcomes by connecting the data from national breast screening programmes to data on cancer cases and long-term survival from national cancer registration systems. And we can only fully understand the impact of ill health on employment status and economic activity by connecting data from health records to data on earnings. Such understanding is crucial to inform policies to improve the current situation, one where economic inactivity due to long-term sickness in the UK has reached a record high of 2.8 million people, representing a key risk to the economy, the government's fiscal position and the NHS.

However, far from being routine, successful linkages of different data sources are all too uncommon, especially those that bring together data from the NHS with data from other sectors, such as census, education or earnings. In England, barriers to such cross-sectoral linkages include the lack of streamlined data-sharing processes between NHS England and the Office for National Statistics (ONS), and lack of clarity about how to comply with the common law duty of confidentiality when linking health and non-health data.

Safe and secure data access

Everyone's health records contain sensitive information that is personal and private. There is a human story behind each data point, and the privacy, confidentiality and security of health data must be taken extremely seriously by all.

The internationally accepted 'Five Safes Framework' (**safe data, safe research, safe people, safe settings, safe outputs**) is widely used to guide research and analysis using health data. Designed by UK experts, the framework protects the privacy and security of peoples' data, ensures that data are used for the public good, and guards against misuse.

To comply with this framework, data custodian organisations:

- de-identify (remove any information such as NHS number that could directly identify a patient) or completely anonymise data (**safe data**) wherever possible;
- make data available only for approved uses for public benefit (**safe research**);
- make data available only to appropriately trained, certified and authenticated analysts (**safe people**);
- provide data wherever possible within highly secure computing environments called secure data environments (SDEs) (**safe settings**). SDEs operate like a reading library rather than a lending library, in that analysts cannot download or export any person-level data and must leave them where they are;
- check the results of any analysis (for example tables or figures) before they are exported from SDEs, to ensure that they could not be used to identify any individual (**safe outputs**).

Recent years have seen a network of SDEs develop across the four nations of the UK, providing analysts with secure remote access to de-identified health data in a protected environment.

Chapter 5 explains how SDEs across the four nations enable access to national-level NHS datasets, data at regional levels and data from other sectors outside the NHS.

Barriers to using health data for public benefit

Several barriers need to be overcome to enable and encourage more beneficial uses of health data. These barriers and potential solutions are outlined in **Chapter 6**.

- Long-term investment in national health data infrastructure is needed, rather than short-term initiatives with unrealistic timelines for delivery.
- Streamlined processes, economies of scale and reducing unnecessary complexity are essential to make the most of limited resources.
- Data custodian and controller organisations should be rewarded for rapid, efficient and secure access to data and services that improve the productivity of data users in generating public benefit, while maintaining the security of the data.
- Strategic partnerships between the health and care system, government bodies, academic research institutions, major public and charitable funding agencies and the life sciences industry are needed to fill the substantial capacity gaps in information governance and in data management and curation, especially within NHS England.

Recommendations

We make five recommendations for transforming the national health data ecosystem and overcoming these barriers. Although these focus on England, reflecting what the commissioners of this review asked for, their principles apply across all four UK nations. The recommendations are explained in full in **Chapter 7**.

Above all, we need to recognise our national health data for what they are: critical national infrastructure that can underpin the health of the nation. They should be treated as such with a strategy, leadership and investment to match.

We also recommend the establishment of a national health data service for England, embedded within existing organisational structures but with accountable senior leadership and a ring-fenced budget. Its main purpose would be to oversee a service to support streamlined, secure research and analysis of health data by approved analysts. It would establish a single data access system for datasets held across England and would follow a priority list for enabling access to key NHS data assets, starting with general practice data. It would work with the ONS, the UK Statistics Authority and relevant organisations in the UK's devolved nations to develop, improve and streamline mechanisms for the sharing and linkage of data across sectors (specifically linkage of NHS health data to health-relevant data from other settings) and across UK nations. We note that there may be similarities or overlap in this recommendation with the government's emerging plans for a National Data Library.

None of this can be achieved without the ongoing support of patients, the public and health professionals. Ongoing engagement with and meaningful involvement of these groups is strongly recommended to shape these advances, ensure transparency on how health data is being used and inform a single opt-out system.

In summary, the UK's complex and inefficient data systems prevent and delay crucial analysis of health conditions affecting millions of people across the UK. We are letting patients and their families down and no change is not an option.

We must focus on **making the simple easy** (for example access to national datasets that already exist) **and the difficult possible** (for example linking national NHS data to datasets from beyond the NHS). This will require coordinated action across multiple organisations and stakeholders to ensure the greatest benefits for patients and the public from health-relevant data across the UK.

Getting this right holds a great prize. Efficient, effective and trustworthy access to our rich abundance of health data will lead to a step change in the UK's research and innovation capability, enhance healthcare, health service planning and delivery, and bring significant economic and societal benefit to the whole country.

Key recommendations

Recommendation One

Major national public bodies with responsibility for or interest in health data should agree a coordinated joint strategy to make England's health data a critical national infrastructure

Making NHS data a critical national infrastructure will boost analysis and research to improve health, wellbeing and economic productivity. We recommend that all major national public bodies that generate, collect, manage, curate, fund or use health-related data in England should sign a commitment to: reduce ecosystem complexity; coordinate long-term planning and investment in publicly funded health data infrastructure; support a national health data service; ensure ongoing nationally coordinated engagement with patients, public, health professionals and politicians; a UK-wide strategy for data access and trustworthy governance; a UK-wide system for SDE standards and accreditation.

Recommendation Two

Leading government health and research bodies should establish a national health data service for England with accountable senior leadership

A national health data service will accelerate research that benefits society. NHS England (NHSE), the National Institute for Health and Care Research (NIHR), the Departments of Health and Social Care (DHSC) and Science, Innovation and Technology (DSIT), and UK Research and Innovation (UKRI) should establish this service to support research and analysis using health data, delivered in partnership with academic, NHS and industry-based research users. It should be led by a senior executive director and have a ring-fenced budget and regularly published performance metrics. Its core functions would be to:

- establish and oversee a **single national health data access system** for England;
- lay out a clear roadmap for data services and dataset provision, complementary national and regional data infrastructures, and streamlined, standardised data governance and access;
- work with the devolved nations, the Office of National Statistics and the UK Statistics Authority to deliver secure, efficient, cross-UK and cross-sectoral data sharing, access and linkage;
- implement an acceptable, transparent investment strategy for health data infrastructure and models for data access cost recovery and pricing.

Key data priorities for this national service should be to:

- **establish a national system for general practice data**, enabling secure access to comprehensive, whole-population, structured, coded general practice data, linkable to other data sources and accessible for a wide range of beneficial uses;
- **enhance and accelerate access to other major national and regional NHS data assets**: hospital episodes, medicines data, lab data (including genomics), national audits and registries, screening data and unstructured clinical data (including imaging and free text);
- **transform access to data from other sectors linked to health and care data at national scale**.

Recommendation Three

The Department for Health and Social Care should oversee and commission a strategy for ongoing, coordinated engagement with patients, public, health professionals, policymakers and politicians

The DHSC should commission a coordinated, multi-organisational strategy for ongoing engagement across society. The wide range of potential data uses should be shaped by the input and experience of patients, public and health professionals, while understanding how best to provide transparency of how data is used for all groups. Major areas of emphasis should be better understanding the perspectives of health professionals, especially GPs; accelerating patients' access to their own health data; and informing a single, centralised, national health and care data opt-out system in England, without imposing any burden on busy GPs.

Recommendation Four

The health and social care departments in the four UK nations should set a UK-wide approach for data access processes and proportionate data governance

A UK-wide approach to streamline data access processes and foster proportionate and trustworthy data governance will enable more and better health data analysis and research. The aim should be for trusted researchers and analysts conducting responsible analyses in the public good to be able to rapidly access the de-identified data they need, while ensuring that data cannot be inappropriately accessed. The approach should be set by the health and social care departments of the UK's four nations and developed with patient and public involvement. It should confront legal and regulatory complexity by providing clear guidance on current approaches, proposing improvements that reduce unwarranted variation, and recommending where new or revised legislation is needed.

Recommendation Five

National organisations in the four UK nations should develop a UK-wide system for standards and accreditation of SDEs holding data from the health and care system

The increasing use of SDEs for maintaining greater control over the sensitive health data accessed by approved researchers has been a great advance in recent years. With more SDEs being set up all the time, a UK-wide system for standards and accreditation of SDEs will accelerate the safe use of health data for patient and public benefit. The UK Statistics Authority and health and social care departments in the four UK nations, with input from relevant UK-wide organisations such as Health Data Research UK, Administrative Data Research UK and Data and Analytics Research Environments UK should lead on the development of: a UK-wide accreditation scheme for SDEs holding data from the health and care system; UK-wide SDE standards to improve user experience and to promote positive user behaviours that benefit all users; and UK-wide policy on avoiding an excess of SDEs.

Cathie Sudlow was supported in carrying out her review by a team at Health Data Research UK.

Cite this report as: Sudlow, CLM (2024). Uniting the UK's Health Data: A Huge Opportunity for Society.

https://doi.org/10.5281/zenodo.13353747

Scan to read the full review