abpi
Bringing medicines to *life*

Models of access to health data in the UK

Report for ABPI Patient Advisory Council

April 2022

Contents

Foreword

Health data is valuable because it delivers insights that can underpin patient-centred innovation in biomedical and clinical research which in turn can lead to better and more equal outcomes for patients and communities. Improving timely access to high-quality data can support improvements in diagnostics, the development of more effective medicines with fewer side effects or lighter treatment burdens, better management of health conditions, and speedier access to treatments. In an ideal world, our health data would be integrated into a single longitudinal dataset. But the reality is that our health data is held in multiple and dispersed datasets of a multiplicity of types and formats. This means that accessing data can prove complex, with ethical, governance, practical and legal challenges along the way.

The ethical use of health data stands to benefit all key stakeholders, including regulators, policy makers, researchers, companies, and people and communities. However, the willingness of the public to share their health data cannot be taken for granted, especially when historic instances of data suppression and exploitation without a clear legal or ethical basis remain in people's minds. Effective governance as well as access is therefore key: realising the full potential of health data will depend on building and maintaining public trust and the adoption of principles including public accountability and transparency (including allowing the public to have a say in how their NHS and care data is used).

Yet it is vital that we do not lose sight of the end purpose or benefit. As representatives of patient and community organisations, we see the pain and anxiety caused by ill health and disability every day. We also understand how this burden falls unequally on different communities, often exacerbating existing inequalities. Our organisations often exist for the very purpose of tackling the causes of ill health and for improving the lives of people affected.

We therefore need no persuading that improved processes for the collection, curation, quality assurance, sharing, and application of health data to biomedical and clinical research (and to NHS and care service improvement and development) will produce the positive outcomes we seek for all the people we support.

We therefore welcome this report that showcases good practice within the UK health data landscape. The breadth and diversity of approaches is impressive and highlights principles and practices that can be shared, adapted and adopted – not least in respect to efficiency of access. However, the key to maximising the value of health data requires attention to further changes, such as consistency in timely access, interoperability, transparency in decision-making processes, and the breaking down of data silos held across multiple organisations and formats. We have an obligation to work together in order to realise what is surely a shared objective of improved equality and outcomes for all.

ABPI Patient Advisory Council

Charlotte Augst, Chief Executive Officer of National Voices

Emily Reuben, Chief Executive Officer and co-founder of Duchenne UK (DUK)

Jane Lyons, CEO of Cancer52

Dr Tom Nutt, CEO of Meningitis Now

Nicola Perrin, CEO of the Association of Medical Research Charities

Jayne Spink, Translational Research Director at Prostate Cancer Research

Bob Stevens, CEO of the Society for Mucopolysaccharide Diseases (MPS Society)

Richard Torbett,
Chief Executive, ABPI

The research-based biopharmaceutical industry is built on the continuous discovery, development and delivery of new medicines to patients, thereby improving and prolonging life. At each stage of the research and development cycle, analysis of health datasets can enhance the process.

The story of the discovery and development of a new medicine always begins with a review of how the disease affects individuals and groups of patients over time and the degree to which current treatments (if any) can affect it. Early research will be designed to improve the understanding of the mechanisms of the disease process, and which treatments might reduce the symptoms and slow or prevent disease progression. In the digital age, direct analysis of health datasets may increase the efficiency of this essential understanding.

The potential of the NHS to provide a UK wide health dataset has been well documented; however, so have the challenges to achieving this. The ABPI "Unlocking the promise of UK Health Data" report in 2020 highlighted several issues, including fragmentation across multiple datasets and efficiency of access processes:

'The most significant issue faced by pharmaceutical researchers is the unpredictability and inefficiency of the data-finding and access process in the UK' [1].

There are also challenges identified by patients around industry access to health data, in particular data that does not have specific patient consent. The Understanding Patient Data report in 2020 "Foundations of fairness: where next for NHS health data partnerships?" found four key points:

- All partnerships must aim to improve health and care for everyone
- NHS bodies need support and guidance to negotiate fair terms for agreements with third parties
- Public accountability, good governance and transparency are critical to maintain public confidence
- The public should have a say in how NHS data is used[2]

Over the last few years, a number of initiatives have been under way to improve the transparency and predictability around access to, and use of, health datasets in the UK. This includes the launch of the Health Data Research Innovation Gateway by HDR UK through which a large number of UK datasets are now discoverable. However, it can be challenging for companies and patients to understand what data is available now, on what terms, for what types of research, in which ways, and examples of data research projects are hard to find.

This report aims to highlight recent improvements in the UK health data landscape, to summarise on behalf of the UK biopharmaceutical industry what types of frameworks for access to and analysis of UK health datasets are currently available, and to give case studies of where they have worked well. Trusted Research Environments are introduced as these are likely to become central to the future of efficient access by researchers, whether in the NHS; academia, industry or the charity sector. The report also makes all findings available to the public to increase the transparency of data access models.

If we can improve the efficiency of the process of discovering the most relevant datasets, discussing whether a dataset can be analysed for a particular research purpose or not and then contracting for the analysis to be undertaken, then more data science projects – and investment – can be attracted to the UK to fund improved data collection, analysis and insights. This increased transparency will help patients understand, review and influence health data availability, access and analysis. In turn, these analyses can lead directly to improved patient outcomes; to refinement of patient pathways and a more efficient NHS; and more precise development and delivery of novel medicines for the benefit of UK patients.

Richard Torbett,
Chief Executive, ABPI

Introduction

This report describes the range of current models for UK health data access for pharmaceutical company researchers, with details on major datasets and case studies. As described in the ABPI publication 'Unlocking the promise of UK health data', analysis of health datasets can contribute to each stage of the cycle of research, discovery, development and delivery of new medicines.

The impact of health research

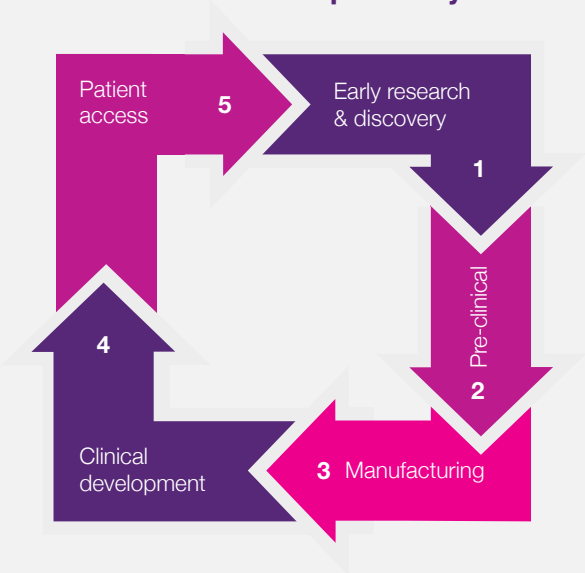
At all stages of the cycle, appropriate health data access for developing and delivering new treatments is a shared priority between patients, policymakers and companies.

It is vital that patients have control over access to their data for research, and that the legal and information governance controls on data use are supplemented with regular dialogue with patients to ensure support for the research and to avoid any surprises. Health data research has the potential to accelerate the development of new treatments, improve healthcare management and delivery, inform health system design, and track disease and health in real time.

This value can be shared between all stakeholders sustainably if the system is transparent, fair and equitable. However industry health data access has been controversial in the UK, with concerns raised about privacy, profit and patient choice. In addition, approved health data access for research has often been unpredictable and inefficient. This report aims to help researchers know what data is currently accessible and how, to help decisions on placing health data science projects in the UK and to increase the transparency of current access models to patients.

The report is based on desk and interview research on UK health data assets that are accessible, or will shortly be accessible, to industry researchers. Health data assets were initially prioritised by likely interest to ABPI members, and then interviewed. Respondents were given the choice of a virtual interview with a researcher, or completing the survey via an online form. Desk research was also conducted to complement the interview and survey data, with some data assets covered in summary form if they are already well known or just starting up, or were not able to fit into the interview schedule.

Health data supports every stage of the medicines development cycle



1 Early research & discovery
- understand disease
- define unmet need
- identify targets
- develop interventions

2 Pre-clinical development
- prove mechanism of action
- assess toxicology/safety

3 Manufacturing, formulation
- develop reproducible, efficient processes and specifications

4 Clinical development
- first-in-human
- clinical pharmacology
- registration studies
- post-approval monitoring

5 Patient access
- MHRA - clinical safety & efficacy
- NICE - HTA
- NHS - pricing & contracting

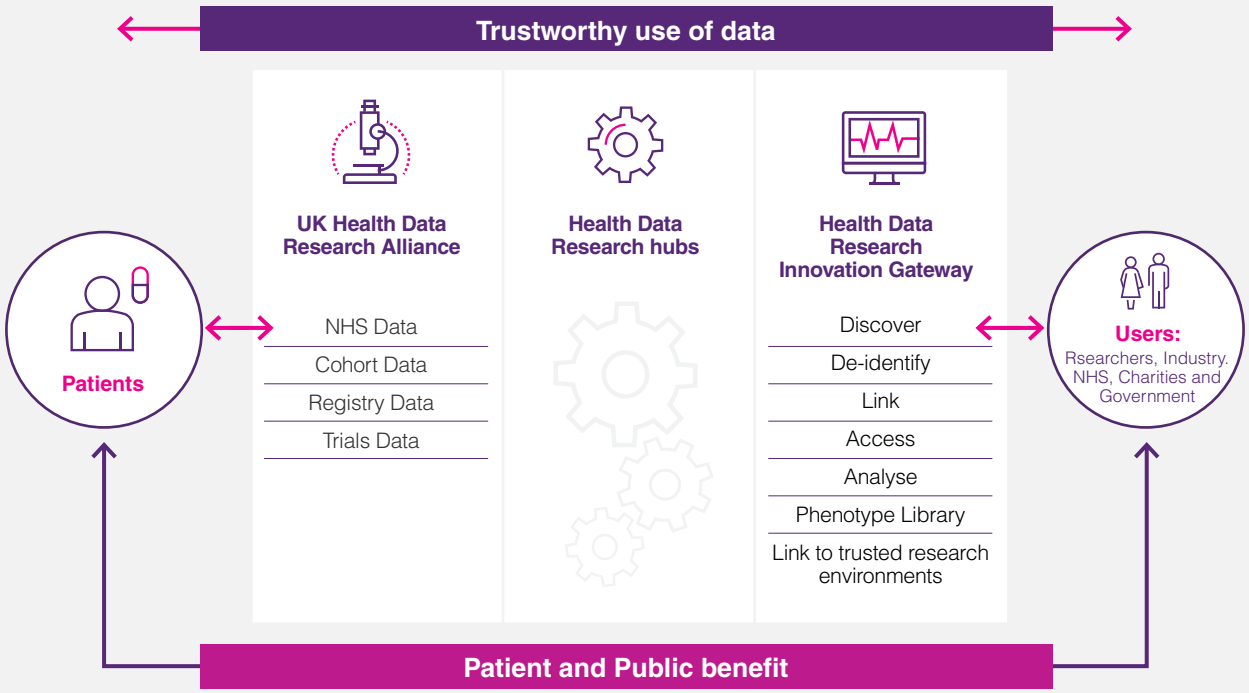
Recent progress

Over the last few years, progress has been made, particularly through the initiatives coordinated by the new national institute HDR UK (Health Data Research UK).

This has done three things:

- It has brought a number of key stakeholders, largely NHS data controllers, together within the Health Data Research Alliance. This currently has over 60 members, committed to work together to '…provide access to rich and diverse health data for research and innovation in a trustworthy and ethical way'.
- It has established the Health Data Research Innovation Gateway, a web portal that '…provides a common entry point for researchers and innovators to discover and request access to UK health datasets for vital research…' This currently provides information on around 650 datasets as well as tools to help researchers analyse them.
- It has established nine Health Data Research Hubs, that can provide 'expertise, tools and knowledge' to help researchers on specific disease areas, on specific datasets, or on specific applications.

Uniting the UK Health Data to make discoveries that improve people's lives



This figure describes the three key parts of HDR UK's use of data model between patients and data users: the Alliance, which brings together data controllers to standardise and increase legal and information governance data access processes, the Hubs, which support data controllers and researchers to find, access and research data in specific areas, and the Gateway, a web portal to make data, algorithms, techniques and processes findable.

Major areas to consider while planning UK health data research

Current access models for pharmaceutical UK health data science

Researchers planning data science using UK health data can approach different organisations depending on their familiarity with the UK data context. Organisations with an understanding of the disease area and available UK datasets for a specific research question will go direct to the data custodian. Those who are less familiar may prefer to explore possible datasets using the Health Data Research UK Gateway[3], which has searchable details of hundreds of accessible datasets. There are also intermediary organisations that can support with data access or deliver an integrated research service: there are many specialist Health Economics and Outcomes Research (HEOR) and data science companies, as well as the larger CROs, and the HDR UK Hubs[4] are able to support in specific areas.

The broad theoretical range of commercial arrangements has been described and debated significantly over recent years, with a workshop by AMS, CASMI and HDR UK early in 2020, describing four major types of arrangement:

- Fee for data access
- Fee for access and service
- Share of revenue (of end product)
- Share of cost saving

In practice, all of the arrangements described below fall into the first two of these categories and pharmaceutical researchers should expect to agree a fee in advance of data access.

UK health data access is well regulated, and researchers are recommended to procure expert advice. The laws impacting data access include the common law duty of confidentiality, the Human Rights Act, and data privacy law such as GDPR, Data Protection Act, etc. The National Data Guardian has set out a Code of Practice and standards for health data access, which is also regulated by the Information Commissioner's Office. Some datasets have been built on specific patient consent including for company research. Others may contain an opt-out or opt-in for company research, meaning a subset is available, and finally there are regulations and organisations that clarify research that can be done on data without consent, such as the Health Research Authority's Confidentiality Advisory Group[5].

Pharmaceutical researchers must be aware that public opinion on health data access starts with a general scepticism about industry access to NHS data, however the use of real-time health data to monitor and manage the COVID-19 pandemic has increased the importance and acceptability of health data access. Patient engagement is vital when considering UK health data research. This patient engagement will be delivered by the data controller to meet their requirements, and it is always worth a researcher considering what additional patient engagement and ongoing transparency can ensure that a research project starts and finishes with patient support. Organisations such as the NHS Centre for Improving Data Collaboration, the NIHR and HDR UK can help with patient engagement.

Health data has been an emerging policy priority for some years, including the Office for Life Sciences Five Principles for health data access to industry researchers, the Department of Health and Social Care Data Security Standards, and the Code of Conduct for Artificial Intelligence and other data-driven technologies. The Goldacre Review was published in April 2022 recommending the modernisation of NHS data analysis and a move towards reproducible analytical pipelines in Trusted Research Environments. The DCMS Data Consultation and NHS Data Strategy are expected to report in 2022.

There are national and regional initiatives and programmes to increase the secure access to UK health data. These include the bodies that process access requests, such as the Data Access Request Service and the Office of Data Release. There are national bodies that grant access, such as the Clinical Practice Research Datalink, NHS Digital for England, and corresponding bodies in the devolved nations. There are national initiatives in key areas, such as UK Biobank, Genomics England, and Our Future Health. Many organisations use trusted research environments (TREs) or are developing TREs, and NHS Digital has developed a TRE to deliver COVID-19 research which has the potential to become a national health TRE.

Summary findings on data access models

We contacted 21 data assets, and every respondent confirmed that data, or insights from it, was or would be available to industry data scientists for health data research. Conclusions from this small qualitative survey must be treated with caution, and the following summary should be considered with this caveat in mind.

Access processes are available, generally via HDR UK Gateway or direct contact

The majority of respondents noted that their approach to industry researchers, their industry access process, and their terms and conditions were publicly available, however these documents were not all easy to find. The HDR UK Gateway "Data Access Request" panel is helpful in showing URL links to access information. For most cases it will be necessary for a researcher to make direct contact with the data controller concerned to fully understand the approach, process, and terms and conditions.

Most datasets recommended that commercial researchers found out details about their dataset from the HDR UK Gateway or directly from the dataset team. Most datasets have metadata available on the HDR UK Gateway, with others making metadata available after signing an NDA or formally registering interest. One fifth stated that they did not distinguish between industry and academic researchers, however their other answers showed that this was focussed on the criteria for granting access, and differences remained in other areas such as pricing.

Encouragingly, almost half of the respondents responded that they had fixed turnaround times for access, and some of these respondents measure and report them as KPIs, though these turnaround times and KPIs were not easily available. Most data sources had dedicated staff working on industry access.

All access must be approved and must benefit patients

All respondents reported overarching usage restrictions, with consistent themes being:

- Showing benefit to the NHS or patients
- Approval by a patient group or representatives
- Approval by a formal data access oversight panel
- Publication of results and acknowledgement of data source

Those datasets whose access restrictions were variable mentioned: bespoke commercial terms, specific data access panel requirements such as scientific rationale, and conditions restricted to specific projects. There is not a standard definition of “approved” or “benefit patients” so each data controller may interpret and implement this differently.

Access models are moving towards trusted research environments

A range of access models was possible, with some datasets allowing more than one model, and most using or moving towards Trusted Research Environments, which is strongly justified and recommended by the Goldacre Review. Access models included working on a copy or anonymised extract of the data, access via a trusted research environment, or research delivered by the data controller or a specified third party. Federated analysis was possible in a minority of datasets at this point.

Row-level data analysis was possible in most of the datasets, and for those where this is not possible, the data controllers can do this analysis themselves on behalf of the researcher. Those datasets which could enable recruitment of patients to trials always did this via referral to clinical teams, with no dataset suggesting direct contact was possible.

Fees for services are universal, with no examples of conditional fees found

Most datasets had different charging structures for industrial users compared to academic or NHS users: NHS Digital was an outlier in not changing pricing by researcher type. Commercial pricing is bespoke for the majority of data sources, with the themes of cost recovery and returning value to the NHS coming up repeatedly.

This research showed that the larger datasets have not yet made agreements around a share of future value, and fee for services is the norm. All respondents charged service fees, most charged data access fees, while a minority would consider data access agreements based on a share of future value and a smaller group would consider a share of resulting cost savings. However, as we did not find any examples of share of future value or cost savings, this is reported as an intent rather than a reality. For all respondents, IP developed in the research belongs to the researcher, is negotiable, or can be shared or licensed.

Trusted research environments

Innovations in information technology have allowed a move from the data being sent to the researcher, to the researcher logging into the data.

In the former model, a copy of data is sent to the researcher who can only use it following specific project permissions and general legal and regulatory restrictions.

It creates a risk that the copy of the data is hacked, leaked or otherwise illegally accessed, or that the researcher does not comply with the permissions or regulations. In the latter model, the researcher is granted access to a computing environment where they can do research on the specific project data but not export it. This avoids making copies of the data and allows the data controller to control and monitor any use of the data.The latter model is now largely delivered through "Trusted Research Environments" or TREs, which are software and hardware platforms that enable data controllers to control and grant access to researchers, to restrict the research that can be done, and to decide on the export of any data or analytic outputs.

The national policy on the use of TREs for accessing NHS data is expected in 2023 as part of the NHS Data Strategy, and is strongly justified and recommended by the Goldacre Review. The TRE model is not without challenges as the TRE processes and functionality must find a balance between the protection of data with the enabling of research. For example, in some cases NHS Digital will not allow researchers based outside the UK to access its data via the TRE, which could be a limiting factor for industry researchers with a therapy area team based outside the UK.

TREs are technically complicated and rely on a set operational and policy decisions that need to be set and then explained to research users and patients. For example, the TRE may not offer the analytical tools that researchers are used to, for reasons of privacy protection, cost, or computer power. Researchers who are used to combining an extracted dataset with their own may find this more challenging in a TRE. If neither of the two datasets can travel, federated analysis will be required and some TREs may not permit federated analysis across multiple datasets.

The HDR UK alliance has recently completed a paper on Building Trusted Research Environments,[6] highlighting the examples of the Scotland Data Safe Haven programme, the UK Secure eResearch Platform in Wales, the Genomics England Research Environment, the Office of National Statistics Secure Research Service, the UK Data Service Secure Lab, the NHS Digital TRE for England and OpenSAFELY. It notes the broad stakeholder support of TREs, and gives a statement of agreed guiding principles for governance and instances of best practice for the operation of TREs, based around the Five Safes. UK health data researchers should expect continued movement towards the TRE model of access, most likely based on the principles described in this report. However the functionality, terms and potential for linking datasets within and between TREs are not standardised and it will be important to identify good practice and lessons learned as TREs are used.

Summary of data sources covered in this report

Name	Type	Size	Focus	Access model
Akrivia Health	Priv	3m	Psychiatric	Access via TRE
Alleviate	Hub	17k	Pain & arthritis	In set-up
Breathe	Hub	9m+	Respiratory	Extract/TRE/ commissioned
Cancer Research UK	NFP	Depends on study	Cancer	Variable
CPRD	Gov	60m, 16m active	Primary care	Extract/TRE/ commissioned
DATA-CAN	Hub	5m population	Cancer	Commissioned research
DataLoch	Gov	1.5m	SE Scotland	Access via TRE
NHS Digital	Gov	55m population	England	Extract/TRE
NHS Digitrial	Gov	55m population	Trial ops	Commissioned service
Discover NOW	Hub	2.3m population	NW London	Access via TRE
eDRIS	Gov	4m population	Scotland	Only approved institutions
Genomics England	Gov	c.100k	WGS	Access via TRE
Gut Reaction	Hub	34k	IBD	Access via TRE
Insight	Hub	>500k	Ophthalmology	Extract/TRE/ commissioned
KID/KeRNEL	Gov	1.5m	Kent	TRE in set-up
Our Future Health	NFP	Planning 5m	All disease	TRE in set-up
Pioneer	Hub	6m population	Acute/unplanned	Extract/TRE/ commissioned
SAIL	Gov	3m	Wales	TRE/commissioned
Sensyne Health	Priv	12m (UK)	Hospital care	Commissioned research
THIN	Priv	68m	Primary care	Extract

Click on a name to view case study

Types:

Priv	**Private company**
NFP	**Charity** (OFH has government and private funding)
Gov	**Government funded initiative** (may generate revenue)
Hub	**HDR UK hub** (also government-funded)

Access models and case studies of health data resources

Akrivia Health

akriviahealth.com

Data focus: Neuroscience (mental health and dementia) cohort dataset
Records: 3 million de-identified psychiatric clinical records
Data Controller(s): NHS Trusts and the University of Oxford
Status: Operational

Akrivia's commercial data access process:	
Details	https://akriviahealth.com/data/ or contact@akriviahealth.com
Metadata available?	No metadata published
Usage restrictions?	Usage restrictions are dependent on the type of researcher applying to use the 'Analytics Platform'
Is it possible to link data?	Yes
Are patients recruitable?	Yes
Access model	Access or commissioned research via the TRE, with federated research possible
TRE	Akrivia operate its own TRE

Akrivia's aim is to "enable important insights from mental health patient data". The data within the platform is de-identified patient data contained in end-to-end clinical pathways and cover each interaction and intervention that takes place between a patient and their clinician.

Data is available for all mental illness and dementia domains and can be curated from both the structured and unstructured data fields. Information is from multiple Electronic Patent Record sources and comprises clinical notes, patient histories, admission and discharge documents and clinical assessments.

Alleviate

www.HDR UK.ac.uk/helping-with-health-data/health-data-research-hubs/alleviate/

Data focus: Pain and arthritis cohort datasets
Records: 16.6k records across 13 cohorts on the Gateway
Data Controller(s): Varies by cohort (e.g. UCL, Bath, Imperial, Oxford etc)
Status: Launched in June 21, in set-up

Alleviate's commercial data access process:	
Details	Not published
Metadata available?	For some cohorts, via HDR UK Gateway
Usage restrictions?	Research in the public interest Data can't leave the TRE Users must do data governance training
Is it possible to link data?	Yes
Are patients recruitable?	Yes, for some cohorts
Access model	In set-up: commissioned or direct research in TRE
TRE	Dundee Health Informatics Centre & Scottish Regional Safe Haven

Alleviate is The Advanced Pain Discovery Platform (APDP) Data Hub. The APDP Data Hub is transforming UK pain datasets to be Findable, Accessible, Interoperable and Reusable (FAIR) and is providing expert data engineering to enhance responsible, timely and trustworthy analysis by researchers and innovators, supported by UKRI, Lilly, Versus Arthritis, Health Data Research UK and others.

The four APDP multidisciplinary consortia are:

- Partnership for Assessment and Investigation of Neuropathic Pain: Studies Tracking Outcomes, Risks and Mechanisms (PAINSTORM)
- Psychosocial mechanisms of chronic pain: targets, translation, and therapeutic innovation
- Consortium Against Pain InEquality (CAPE) – The impact of adverse childhood experiences on chronic pain and responses to treatment
- ADVANTAGE visceral pain consortium: Advanced Discovery of Visceral Analgesics by Neuroimmune Targets and the Genetics of Extreme human phenotypes

Once fully operational, Alleviate will be updating its website pages and portal with further information regarding its commercial access protocols and procedures.

Breathe

www.ed.ac.uk/usher/breathe

Data focus: Respiratory conditions cohort and population data, UK & intn'l
Records: 74 datasets with up to 9m records each
Data Controller(s): Cystic Fibrosis Trust, Breathe, SAIL (COVID-19 & Wales pop. data)
Status: Operational

BREATHE's commercial data access process:	
Details	Via email to monica.fletcher@ed.ac.uk
Metadata available?	On request
Usage restrictions?	All data access must be for public benefit A clear scientific rationale must be displayed rather than the sole means of access being financially driven For data held in SAIL, data must be accessed via the SAIL TRE
Is it possible to link data?	Yes
Are patients recruitable?	Yes, for some cohorts
Access model	Many: data extract, TRE, federated, commissioned
TRE	SAIL databank: SeRP, Swansea

Breathe is an HDR UK hub working via its Industry Forum to bring together representatives from multinational companies, SMEs and university spin-outs to raise issues and bring a commercial perspective to the Hub's services. Novartis, which also represents the Industry Forum, is on the BREATHE Executive.

BREATHE's technical host is SAIL (Swansea University), which is experienced in data access. The commercial offering is described on the website, and pricing is cost-recovery plus for commercial clients. IP stays with the researchers requesting data access.

Cancer Research UK

commercial.cancerresearchuk.org/

Data focus: Cancer research cohort datasets
Records: On request
Data Controller: Cancer Research UK
Status: Operational

CRUK's commercial data access process:	
Details	Via its Commercial Partnerships Team commercial@cancer.org.uk
Metadata available?	On request
Usage restrictions?	Usual data safeguarding constraints: research should be continuable by academia/public sector if desired
Is it possible to link data?	No
Are patients recruitable?	No
Access model	Varies by underlying dataset but requires a fair return to CRUK & data generators
TRE	CRUK has a TRE, but it is not a condition of access for most datasets

Cancer Research UK is a charitable research funder. The majority of the funded research gives full or partial data rights to CRUK, and these are managed by the CRUK Commercial Partnerships team. CRUK is currently building a portfolio of commercially available datasets and increasingly ensuring that access is possible for legitimate academic and commercial researchers who can accelerate the development of diagnostics and treatments.

An example of Cancer Research UK's data access is Optimam, https://news.cancerresearchuk.org/2021/09/08/speaking-a-thousand-words-how-a-cancer-image-collection-is-set-to-improve-ai-diagnosis/.

Clinical Practice Research Datalink (CPRD)

www.cprd.com/

Data focus: Primary care population dataset (all conditions)
Records: 60m records. 16m active patients
Data Controller: DHSC
Status: Operational

CPRD's commercial data access process:	
Details	Via www.cprd.com/Data-access or enquiries@cprd.com
Metadata available?	Yes
Usage restrictions?	Published conditions include an assessment of public health benefit, ethics, feasibility and risk to confidentiality
Is it possible to link data?	Yes, via CPRD services
Are patients recruitable?	Yes, via CPRD services
Access model	Extracted data, commissioned research, moving to TRE
TRE	Developing a TRE currently

The Medicines and Healthcare products Regulatory Agency (MHRA) and the National Institute for Health Research (NIHR) jointly sponsor the Clinical Practice Research Datalink (CPRD). Data held by CPRD covers primary care for 60 million patients, including 16 million currently registered patients. The primary care data are near real time, longitudinal and representative of the UK population and can be linked to a range of health-related datasets.

As a not-for-profit, cost recovery UK government research service, CPRD must recoup the cost of delivering research services from data access licence fees. Data access licensing options include access to CPRD data provided by a multi-study annual licence (£330k for large companies) or on an individual dataset study licence basis (£60k for large companies).

Case studies:
https://www.cprd.com/CPRDcasestudies

DATA-CAN

www.data-can.org.uk

Data focus: Oncology population and cohort datasets
Records: 5m population, c.450k in cohorts
Data Controller(s): Hospitals and universities
Status: Fully functional, TRE not yet commercial accessible

DATA-CAN's commercial data access process:	
Details	https://www.data-can.org.uk/health-data/working-with-life-science-companies/
Metadata available?	Yes (HDR UK Gateway)
Usage restrictions?	PPIE group must approve all data access Population research must be delivered by NHS/Uni staff National data not currently accessible to companies
Is it possible to link data?	No
Are patients recruitable?	Only with new REC approval
Access model	Commissioned research
TRE	Yes (currently only for non-profit COVID-19-related research)

DATA-CAN is the HDR UK hub for cancer data access. It is not a data controller but organises access with data controllers as well as delivers data curation and analytical services. Commercial access is possible for most of the datasets.

The population dataset (hospital records covering 5m population) is accessible solely for commissioned research delivered by NHS/University staff behind the firewall, with anonymised research conclusions made available.

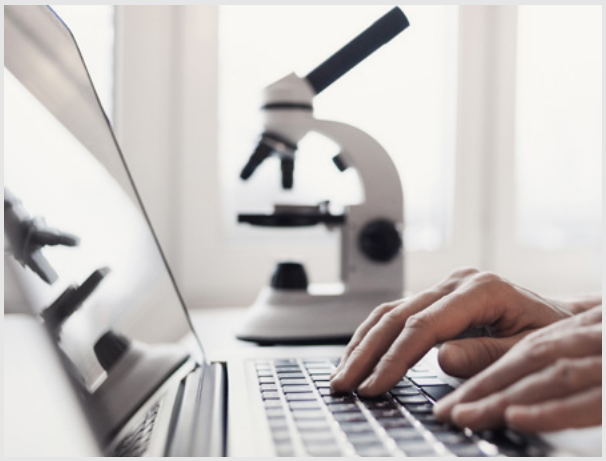
DATA-CAN is currently evaluating the national Cancer TRE which contains all relevant NHS Digital cancer datasets (HES, NCRAS, etc) however this is not yet accessible for commercial research.

DATA-CAN case study: Early Triple Negative Breast Cancer study with Roche

DATA-CAN and Roche worked with the Leeds and Edinburgh Cancer centres to collect, curate and analyse historic data from patients with early-stage triple negative breast cancer (eTNBC), to better understand the treatment and outcomes of NHS patients. The team created a single data model to allow these two datasets to be combined and analysed in a secure environment by accredited NHS and academic staff, who then presented and discussed the results with Roche scientists.

Cancer patients from DATA-CAN's patient and public involvement and engagement group have been involved at all stages of the project and have given their approval to the work. All data remains the property of the NHS, and Roche only saw aggregated results, not anything that can be used to identify an individual patient.

Patients will benefit from the improved evidence to inform the NHS's adoption of new drugs to treat patients with triple negative breast cancer. Patients remain in control of their data both via the NHS national data opt-out, and via the patient engagement and governance of the project itself. The NHS will benefit from an improved understanding of how patients with early-stage triple negative breast cancer are treated in two major NHS cancer centres. It will also benefit from the improvement of data quality for ongoing patient management and any future research. Roche will benefit from an improved understanding of how this group of breast cancer patients is treated in the UK, with the data being used to support drug development and patient access to medicines.



DataLoch

www.ed.ac.uk/usher/data-driven-innovation/dataloch

Data focus: Population, cross-disease
Records: >1.5m records
Data Controller(s): Scottish regional councils and SE Scotland Health Boards
Status: Not yet operational

Key survey responses on DataLoch's commercial data access process:	
Details	https://dataloch.org/
Metadata available?	Yes (register a research interest directly first)
Usage restrictions?	see: https://dataloch.org/data/how-to-apply In particular: Project must demonstrate benefit to NHS and other patients Research must be published
Is it possible to link data?	Yes
Are patients recruitable?	Sometimes
Access model	Not yet operational
TRE	using eDRIS's TRE, DataLoch is also building its own, which will permit researchers running commercially sensitive code in privacy

DataLoch has a comprehensive population dataset of South East Scotland, including social services data which can be linked as appropriate. It is aware of industry interest and keen to offer a full service through to clinical trials, including commissioned research and access via a TRE.

NHS Digital including Digitrials

https://digital.nhs.uk/services/nhs-digitrials

Data focus: Primary and secondary population datasets
Records: 55m population (England)
Data Controllers: NHS Digital
Status: Operational

NHS Digital and Digitrials commercial data access process:	
Details	https://digital.nhs.uk/services/nhs-digitrials and https://digital.nhs.uk/services/data-access-request-service-dars/data-access-request-service-dars-process
Metadata available?	Yes, via HDR UK Gateway
Usage restrictions?	Data security obligations, researcher training in data guardianship Public transparency of research requests, research output must be of public benefit and proportionate to the data requested
Is it possible to link data?	Yes
Are patients recruitable?	Yes, with additional restrictions
Access model	Extracted dataset, or TRE, or commissioned
TRE	TRE available

NHS Digital has responsibility for standardising, collecting and publishing data and information from across the health and social care system in England. NHS Digital holds the largest and most commonly-used English national NHS datasets, including the administrative datasets HES and SUS, COVID-19, data on imaging, emergency care, mental health, workforce, primary care prescriptions, with a full list at https://digital.nhs.uk/services/data-access-request-service-dars/dars-products-and-services. These data are available through applying to the Data Access Request Service (DARS).

In October 2021, NHS Digital took on responsibility for the management of the National Disease Registration Service (NDRS), a collection of data on all cancers, rare diseases and congenital anomalies diagnosed each year in England. Access to NDRS datasets is via the Office of Data Release,[7] which is currently on pause to focus on the handover of data applications to the new system.

Some datasets can be shared with researchers for them to do permitted research, and NHS Digital is moving to all data access happening through trusted research environments.

NHS Digital recommends that industry researchers consider the following points in the process of applying for access:

1. Talk to us early – so the application complies with requirements
2. Have a dataflow mapped out in advance
3. Be clear on information governance responsibilities and roles (data processor/controller etc)
4. Researchers often apply with a pre-formed cohort without knowledge of regulations and rules around access: they must understand the data information security, IG, legal basis and purpose also
5. Combining a robust scientific and IG approach will enable a predictable process of application
6. Understand the basic guidelines around use of data and legal basis that are likely to be necessary

NHS Digital delivers commissioned research, although this is mainly for NHS England and no commercial commissioned research has been done and capacity is constrained. The disease registries have done more commissioned research, including assessments of drugs on managed access programmes for NICE assessment. Third parties can be commissioned to do research services, including NHS Digitrials.

NHS Digitrials is a new organisation set up to deliver a range of data services to support clinical trials and those conducting wider healthcare research. NHS DigiTrials supports delivery of data services to support clinical trials, such as invitations to recruit and communications to trial participants, and has a model that is paid for at cost via licence and service fees depending on complexity of requests.

Researchers can learn more about access via completion of their 'Feasibility Request Form', a link to which can be found here:

https://nhs-prod.global.ssl.fastly.net/binaries/content/assets/website-assets/services/nhs-digitrials/nhs-digitrials-feasibility-request-form

Digitrials case study:
https://www.youtube.com/watch?v=mb5oSMyzNrM

Discover-NOW

imperialcollegehealthpartners.com/discover-now/

Data focus: Population datasets
Records: 2.3m population (North London)
Data Controller(s): North London NHS data controllers
Status: Operational

Discover-NOW's commercial data access process:	
Details	https://imperialcollegehealthpartners.com/discover-now/
Metadata available?	Yes: HDR UK Gateway
Usage restrictions?	Must use TRE Researchers must be vetted Existence of commercial contracts published Research results shared with Discover-NOW Research contracts for 6 months each
Is it possible to link data?	Yes
Are patients recruitable?	Yes: System in place to facilitate trials
Access model	Currently commissioned research, moving to TRE
TRE	Yes

Discover-NOW is the HDR UK hub for Real World Evidence, based on safe, secure access to a population of 2.3 million records in a trusted research environment. The Hub also offers regulation-ready and GDPR-compliant tools, technologies and expertise to support RWE health research and improvements. These include a high-performance analytics environment, methods including machine learning, a feasibility and research recruitment tool, as well as a consent-to-contact register of more than 8,000 people.

The work of Discover-NOW is initially focussed on how real-world evidence can improve understanding of many long-term conditions. This initial research is in partnership with AstraZeneca, one of the Hub's largest contributors, and is focussing on finding new solutions through real-world evidence to help people to manage their Type 2 Diabetes. Discover-NOW does not share copies of data; the data must be analysed in-house or via the TRE. The preferred model for commissioned research is via ICHP, Imperial College, or led by clinical academics in North West London.

Discover-NOW case study: Atrial Fibrillation Algorithm Validation with BMS

Bristol Myers Squibb worked with the Discover-NOW team at Imperial College Health Partners to evaluate the ability of a machine learning algorithm to identify patients at high risk of atrial fibrillation in primary care, using de-identified primary care data from 2.5m patients between 2001 to 2016. The results of this work were published in the European Journal of Preventive Cardiology in June 2020 and another article on budget impact analysis is in process.

The algorithm had been developed using the CPRD primary care dataset. The Discover registry does not cross over with the CPRD dataset, so it could be used for an external validation. BMS delegated the analysis to the ICHP team, who delivered the analysis under the Discover overall Research Ethics Committee approval and after specific approval of the project by the Data Research Access Group. The BMS team found it easy to work with ICHP, with the team able to augment the original analysis with a number of options, and with experience in writing publications based on real-world data. The quality of the work was good, passing peer review scrutiny quickly.

Atrial fibrillation increases the risk of stroke and mortality. 30% of atrial fibrillation is undiagnosed and early detection should improve outcomes, however detection is a fine balance between patient burden, healthcare costs, and diagnostic accuracy. The algorithm performed similarly to development which means it could be useful to narrow the population to detect those at highest risk of atrial fibrillation, who can be cost-effectively screened using ECG. This project shows how industrial health data research can provide insights that are valuable to patients, the health system, primary care doctors, and researchers developing treatments.

eDRIS

www.ed.ac.uk/edinburgh-international-data-facility/updates-events/electronic-data-research-and-innovation-service

Data focus: Population, cross-disease
Records: >4m records
Data Controller(s): Public Health Scotland and Scottish Regional Health Boards
Status: Operational

eDRIS' commercial data access process:	
Details	https://www.isdscotland.org/Products-and-Services/eDRIS/_docs/2019-07-31-eDRIS-Customer-Contract-Terms-v-1.pdf
Metadata available?	Yes: HDR UK gateway
Usage restrictions?	Must prove public benefit Must retain data in the UK Data governance determined by individual data controllers Individual researchers must have approved information governance training Only researchers from "Approved Institutions" allowed direct access to data
Is it possible to link data?	Yes
Are patients recruitable?	Yes: Scottish Health Boards clinicians, via eDRIS
Access model	Commissioned research
TRE	Yes

Since 2013, the Electronic Data Research and Innovation Service (eDRIS) research portal has been a national "Safe Haven," a data hub that provides a state-of-the-art technical facility with high end performance computing.

eDRIS provides a research portal for data held collectively by Public Health Scotland. Currently it only actually hosts COVID-19 data, but other records are available through it.

Industry researchers cannot directly access data, they can commission, or collaborate with, third party researchers (academic/NHS) from "Approved Institutions" who are allowed to use eDRIS data.

eDRIS case study

GlaxoSmithKline: Identifying risk factors for progression to critical care admission and death among individuals with acute pancreatitis: a record linkage analysis of Scottish healthcare databases - https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4916584/

Strategen and Abbvie: Association between respiratory syncytial virus hospitalisation in infancy and childhood asthma - https://onlinelibrary.wiley.com/doi/full/10.1002/ppul.24676

Genomics England

https://www.genomicsengland.co.uk/research/partnerships/discovery-forum

Data focus: Consented cohort including rare genetic disease, cancer, C-19
Records: >17k cancer, >70k rare disease, <35k C-19
Data Controller: Genomics England
Status: Operational

Genomic England's commercial data access process:	
Details	https://www.genomicsengland.co.uk/ or via email to commercialpartnerships@genomicsengland.co.uk
Metadata available?	Yes
Usage restrictions?	All access is via the TRE IG training must have been completed in order to authorise user access Only summary level data (not participant identifiable raw data) can be extracted/ distributed outside of research environment Any information which is extracted must come through the 'airlock'
Is it possible to link data?	No
Are patients recruitable?	Yes (recruitable via treating clinicians)
Access model	Direct access to TRE or indirect via Genomics England Bioinformatics consultancy service
TRE	On premises HPC RE: RE1.0 Cloud based TRE; CloudRE via Lifebit

Genomics England aims to bring benefit to patients using whole genome sequencing, to create an ethical and transparent programme based on consent, to enable new scientific discovery and medical insights, and to kickstart the development of the UK genomics industry.

It was originally set up to deliver the 100,000 Genomes Project; with the completion of the project, Genomics England now provides whole genome sequencing to the NHSE Genomic Medicine Service (GMS), and is undertaking further government-funded initiatives such as the Newborn Sequencing programme. Genomics England grants access to linked genetic and clinical data to approved academic and industry researchers via its trusted research environment.

Genomics England works with industry through its Discovery Forum. In order for industry partners to get access to environment/data, they need to first be approved (Company and initial Research proposal) by the Access Review Committee. All members of the Forum are encouraged to publish all findings and research at the point at which intellectual property for any product is protected, in common with best practice in the pharmaceutical industry.

Data records include whole genome sequences and linked clinical data (including HES, SACT etc) from

- >17k cancer (paired somatic and germline WGS),
- >70k rare disease (probands and relatives where possible),
- <35k COVID-19 patients (mild- and severely-affected).

Industry researchers can commission the Genomics England Bioinformatics consulting team or specific third-party researchers to deliver research on their behalf.

The process for industry researchers to apply for membership to the Discovery Forum, to access the data and resources at Genomics England, is via the Partnership Development team: commercialpartnerships@genomicsengland.co.uk

Gut Reaction

www.gut-reaction.org/

Data focus: Inflammatory bowel disease cohort
Records: 34k
Data Controller: Cambridge University Hospitals NHSFT
Status: Operational

Gut Reaction's commercial data access process:	
Details	https://gut-reaction.org/ibd-data-gut-reaction/accessing-data/industry-requests/
Metadata available?	Yes, via HDR UK Gateway
Usage restrictions?	Benefit to patient health Must have a lay summary of research Must have a clear aim and specify datasets Data must only be used for the purpose specified Required acknowledgment in publications Confidentiality No attempt to directly contact participants
Is it possible to link data?	Yes
Are patients recruitable?	Yes, via the NIHR BioResource
Access model	Direct or commissioned from third party, via TRE
TRE	Owned by Gut Reaction: isolated 'bubble' for each research project (not shared), airlocked.

Gut Reaction is the HDR UK hub for inflammatory bowel disease, based on the NIHR Bioresource for IBD.

Companies access the data in the TRE. Gut reaction can assist with research but does not deliver research.

Clinical study cohort selection and recruitment is possible, through NIHR Bioresource and the **Cohort identification tool** on the HDR UK Gateway.

Insight

https://www.insight.hdrhub.org/

Data focus: Ophthalmology cohort
Records: over 500k across 11 datasets
Data Controller(s): Moorfields and Birmingham hospitals
Status: Operational

Insight's commercial data access process:	
Details	https://www.insight.hdrhub.org/researcher-area
Metadata available?	Yes, via HDR UK Gateway
Usage restrictions?	Value must be demonstrated to have returned to NHS, patients and public, following the "Five-Safes" principles
Is it possible to link data?	No
Are patients recruitable?	No
Access model	Extracted dataset, or TRE, or commissioned research
TRE	uses TRE of University Hospitals Birmingham

Insight is an NHS-led partnership with a mission to improve healthcare by making it simpler for researchers to use large, anonymised sets of patient data in a safe and ethical way. The focus is on eye diseases such as age-related macular degeneration and glaucoma, alongside early detection of dementia, diabetes and heart disease using eye-related datasets. Revenue is licence and service fee based.

Researchers can explore the data access process further by visiting the link noted above, contacting directly via https://www.insight.hdrhub.org/contact, or reviewing the full process at

https://1ac9575b-f9b8-45b9-8ec4-e84cbe9314b6.filesusr.com/ugd/714db5_b44998863a0f4751a551f5c91d39b299.pdf

KID and KeRNEL (Kent & Medway)

https://www.kmkernel.org

Data focus: Population-based primary, secondary and social care
Records: 1.5m records
Data Controller(s): Healthcare Trusts, Primary care entities (CCGs) and Social Services entities in Kent, UK
Status: Data access requests will be reviewed on a case-by-case basis

KID/KeRNEL commercial data access process:	
Details	https://www.kmkernel.org/resources/de-identified-data-access-request-form
Metadata available?	Embedded in request form for KID
Usage restrictions?	Must support health and social care objectives, particularly to Kent/ Medway area
Is it possible to link data?	Yes
Are patients recruitable?	No
Access model	Case by case
TRE	Will operate its own TRE (under construction)

KID and KeRNEL have been designed from the outset to link healthcare and social services/ local authority data, (KID was originally for public sector resource planning research).

KID is still available for historical population research access; KeRNEL is intended to replace it, offering much closer to real-time data as well as analytical and health economic services.

There is enthusiasm for commercial work, and an understanding of commercial imperatives, but obtaining necessary agreements with different data controllers, has made KeRNEL construction take time.

The datasets within the KID are linked using the NHS number as a common patient identifier then de-personalised. All data is hosted and analysed within the Health & Social Care Network (HSCN) meaning that analysis of this de-personalised record level data must be undertaken by analysts with permitted access to this NHS-hosted safe haven.

The Health Economics Unit liaises with the Kent data team to facilitate such access, either by helping researchers to refine their requests or by checking whether the dataset is likely to contain the information they need before they make an application for access.

Our Future Health

https://ourfuturehealth.org.uk/

Data focus: Large general population cohort
Records: In construction
Status: Our Future Health is currently in construction

Our Future Health's data access process:	
Details	https://ourfuturehealth.org.uk/research-programme/ Alternatively, making direct contact info@ourfuturehealth.org.uk
Metadata available?	Once fully operational
Usage restrictions?	Scientific utility is assessed
Is it possible to link data?	Yes
Are patients recruitable?	Yes
Access model	In development
TRE	Yes

Our Future Health is a not-for-profit, charitable company founded in 2020 to develop the UK's largest health research programme, designed to enable the discovery and testing of more effective approaches to prevention, earlier detection and treatment of diseases.

(https://ourfuturehealth.org.uk/about-us/)

Our Future Health will collect and link multiple sources of health and health-relevant information, including genetic data, across a cohort of 5 million people that truly reflects the UK population.

This will create a world-leading resource for academic and commercial researchers to undertake discovery research on early indicators of disease, plus the opportunity to re-contact participants on a risk-stratified basis for secondary studies.

More information on their current programme can be found here:
https://ourfuturehealth.org.uk/research-programme/

All IP will belong to the researcher. Recruitment into the cohort at scale will begin during 2022 and researchers will be able to apply for access to data and samples from 2023.

PIONEER

https://www.pioneerdatahub.co.uk/

Data focus: Population dataset in acute/unplanned care and physiology
Records: Specific cohorts (c.10-150k) from a population of 6m
Data Controller: UHB
Status: Operational

Pioneer's commercial data access process:	
Details	https://www.pioneerdatahub.co.uk/about/working-with-us/ https://www.pioneerdatahub.co.uk/data/data-request-process/
Metadata available?	Yes, via HDR UK Gateway
Usage restrictions?	Data usage must follow the PIONEER ethics approvals and must be reviewed and assessed by the PIONEER Data Trust Committee
Is it possible to link data?	No
Are patients recruitable?	No
Access model	PIONEER can export an anonymous extract, researchers can use the PIONEER TRE, or PIONEER staff can deliver research
TRE	Microsoft Azure based, fully security tested.

The PIONEER programme is the Health Data Research Hub for Acute Care, led by the University of Birmingham and University Hospitals Birmingham NHS Foundation Trust, in partnership with West Midlands Ambulance Service, the University of Warwick, and Insignia Medical Systems. PIONEER links acute care health data at the individual patient level from community and hospital healthcare providers.

Researchers are able to access via pioneer@uhb.nhs.uk or a request form that is straightforward to complete: https://www.pioneerdatahub.co.uk/data/data-request-form/ and via: https://www.pioneerdatahub.co.uk/wp-content/uploads/PIONEER-DATA-REQUEST-FORM-V4-14-09-2020-ES-approved.docx

This provides, for the first time, a record of the journeys people undertake when they have unplanned health concerns. A dedicated commercial team has been established, and a list of services and cost drivers can be found here:

https://www.pioneerdatahub.co.uk/data/data-services-costs/

Case studies:
https://www.pioneerdatahub.co.uk/project-summaries/

SAIL Databank

saildatabank.com/about-us/overview/

Data focus: Population, all disease, principally NHS Wales datasets
Records: 3m
Data Controller: Swansea University
Status: Operational

SAIL's commercial data access process:	
Details	https://saildatabank.com/application-process/following-approval/#remote-access
Metadata available?	Yes, via the HDR UK gateway
Usage restrictions?	Ultimate public benefit Data volume requested must be proportionate to the research questions being asked All research conducted within SAIL should lead towards impact and outputs e.g. technical reports, data brief whitepapers, academic papers, submissions to NICE, etc. would be expected to be put into the public domain as appropriate Researchers may not use SAIL for performance management, nor league table creation
Is it possible to link data?	Yes
Are patients recruitable?	Yes, via partners
Access model	Commissioned research in TRE preferred, TRE access possible
TRE	Yes. SAIL operates its own TRE.

SAIL is a longstanding data service supplier, which now additionally provides technical resources for a number of other HDR UK hubs, alongside the datasets that it publishes in its own right, including health and population datasets for Wales.

As well as technical resources, it has an analytics team who are assigned to commercial projects during the scoping phase, and various research programme groups who can be collaborated with in terms of the use, expertise and experience of use SAIL and its data.

Sensyne Health plc

www.sensynehealth.com/

Data focus: Population based from NHS and US
Records: 12m in UK, 13m in US
Data Controller(s): NHS Trusts and US medical entities
Status: Operational

Sensyne's commercial data access process:	
Details	https://www.sensynehealth.com/
Metadata available?	No. Bespoke service for researchers, allowing them to assess feasibility of a specific approach to a research question and/or a suitable cohort for analysis.
Usage restrictions?	Research leading to patient benefit is essential. This must be stated in the Information Governance request There are no constraints on use of results
Is it possible to link data?	Anonymised data from different NHS Trusts can be linked
Are patients recruitable?	Not by Sensyne, but if a suitable cohort is identified, Sensyne may be able to refer researchers to relevant clinicians, who can recruit
Access model	Commissioned research
TRE	No: companies design research questions; Sensyne delivers the research on their behalf.

Sensyne has two significant "divisions" of relevance:

- **A product division** for the global healthcare sector. This has two product lines, simplified here:
  - A range of "digital health" applications (running on mobile phone technology), providing patient self-monitoring, including dip tests and diabetes (finger-prick) tests. These are integrated with reporting systems for clinicians.
  - Sense, a cloud-based, commercially-packaged version of the company's AI technology, as a decision support engine in acute medicine.
- **An analytics division**, supplying anonymised data and analytics, at scale, to the healthcare sector and pharmaceutical companies in particular. Sensight, a pre-packaged analytics/data platform (with a significant AI capability), uses a subscription model (£25k/seat/annum).

Sensyne also contracts on a bespoke basis with commercial firms.

THIN (Cegedim Health Data UK)

https://www.cegedim-health-data.com/cegedim-health-data/thin-the-health-improvement-network/

Data focus: Primary care population dataset across diseases
Records: >68m records, 9m current patients
Data Controller: THIN, with data use governed by HRA approval
Status: Operational

THIN's commercial data access process:	
Details	www.cegedim-health-data.com/scientific-research-contact/
Metadata available?	No, but full description of data is available after NDA. Feasibility maybe conducted by Cegedim Rx or by the applicant.
Usage restrictions?	For any publications following research/analysis, also where the outcome of the research is to provide an algorithm or tool, SRC approval will be required The data is not to be used for any insurance purposes and no attempt should be made to re-identify the patients or to be link the data to another data source Data cannot be shared with any third party without prior agreement Data is to be destroyed after the data term/licence expires
Is it possible to link data?	No
Are patients recruitable?	Yes
Access model	Extracted dataset, or TRE, or commissioned
TRE	THIN maintains and operates its own TRE

The Health Improvement Network (THIN) is a large European proprietary database established in 1994 of fully anonymised and non-extrapolated Electronic Health Records, reportedly over 68 million patient files holding patient characteristics and clinical outcome data, collected at the physicians' level.

Cegedim Health Data provides Real World Data and Evidence to enable advancements in patient outcomes. This hub contains unprocessed European patient healthcare data. THIN's website states access to the full longitudinal anonymised patient database with an average of 7 years' history per patient, from patient characteristics to clinical outcomes. THIN has worked with more than 25 pharmaceutical companies.

Other data sources and relevant organisations

Relevant organisations that do not grant access

AIMES

https://aimes.uk

AIMES is a technology provider working with organisations such as Cancer Research UK and Birmingham to set up Trusted Research Environments.

LHCRE: Y&H

https://yhcr.org

The aim of the Local Health and Care Record Examplars programme is to create an information sharing environment that helps our health and care services continually improve the treatments we use, ensures that care is tailored to the needs of each individual, and can empower people to look after themselves better and make informed choices about their own health and care. There are five LHCREs: Greater Manchester, Wessex, One London, Yorkshire and the Humber, Thames Valley and Surrey.

The primary focus of the LHCRE is to create integrated health and care records for individual care. However, the five regions are also considering how shared health and care records could be used to support purposes beyond individual care such as improving health and services through research and planning. The LHCREs do not grant data access or supplying services to industry researchers.

NHSX

https://www.nhsx.nhs.uk

NHSX is part of NHS England, supporting local NHS and care organisations to: digitise their services, connect the health and social care systems through technology, and transform the way patients' care is delivered at home, in the community and in hospital. NHSX contains the Centre for Improving Data Collaboration which supports better data-driven partnerships between the NHS and industry, ensuring these promote patient and public benefit.

Data Access organisations that are not described above

BHF Data Hub

https://www.HDR UK.ac.uk/helping-with-health-data/bhf-data-science-centre/

The BHF Data Science Centre is a partnership between Health Data Research UK (HDR UK) and the British Heart Foundation (BHF). A key goal is to ensure that bona fide researchers UK-wide (including from the NHS, academia and industry) can discover, access and analyse national healthcare data relevant to a wide range of cardiovascular research questions. Started in January 2020, it has initial investment from the BHF for £10 million over the first five years, with the current focus of the BHF Data Science Centre being on the response to COVID-19. It does not have an industry partner on its website.

DATAMIND

https://www.HDR UK.ac.uk/helping-with-health-data/health-data-research-hubs/datamind/

DATAMIND is an HDR UK hub providing innovative data resources for mental health research and innovation, launched at the same time as Alleviate in June 2021. The Hub will improve the discoverability and usability of diverse data sources for research to help improve the lives of people with mental health problems. The DATAMIND Hub will maximise the value of this data by safely and securely bringing together data from diverse sources, including health records, schools and administrative data, charity data, research trials, genomics, longitudinal studies and cohort data.

IQVIA Cancer Data Network

https://www.iqvia.com/locations/united-kingdom/solutions/cancer-data-network

IQVIA is building a Cancer Data Network in the UK. The IQVIA Cancer Data Network enables organisations to improve outcomes for NHS cancer patients across the UK, improving cancer service delivery, matching patients to clinical studies, identifying treatment variations with national reporting and benchmarking, and making high quality data available for research. Details of sites, datasets and access processes will be available when the Network is launched.

NHS Genomic Medicine Service (GMS) Research Collaborative

Data collected as part of routine clinical practice via the NHS Genomic Medicine Service is not available for commercial or academic researchers to access. The NHS Genomic Medicine Service Research Collaborative has been brought together as part of the NHS Long Term Plan aim to support research and development by facilitating genomic research on a national scale.

The NHS GMS Research Collaborative offers two main services:

- An Early Feedback Review Service, in partnership with the NIHR Study Support Service, providing feedback on developing research proposals to inform research planning
- Review of research proposals that are seeking support from the NHS GMS via the Steering Committee

Researchers are invited to submit to each of the above processes via the submission forms on the NHS GMS Research Collaborative web page (see below). By submitting a research proposal, researchers agree to the principles of the NHS GMS Research Collaborative (outlined on the web page). A key principle is that all research and industry collaborations should be for the potential benefit of patients and the outcomes must be reported back to the NHS to support improvement in patient care.

www.england.nhs.uk/genomics/genomic-research/nhs-genomic-medicine-service-research-collaborative/

NIHR Bioresource

https://bioresource.nihr.ac.uk/

NIHR Rare Diseases BioResource has recruited 150,000 affected participants and, in some cases, their relatives, to give samples and consent for data access. It works in more than 50 disease areas, including in immunity, neuroscience, haematology, rheumatology, and cardiovascular disease.

The BioResource is designed to make research and development more efficient, with a single consent, a single repository, and a single database. The samples and data have been developed to accelerate discovery in disease research.

NIHR Biomedical Research Centres

The NIHR BRCs are involved in health data research, and some have specific data access asset and services. An example is Imperial BRC's Clinical Analytics, Research and Evaluation team (iCare), which focusses on two workstreams:

a) routinely collected clinical information for direct patient benefit, clinical analytics and research

b) digital interventions and clinical tools for patient benefit. Metadata and the data access process can be found at https://imperialbrc.nihr.ac.uk/facilities/icare/explore-and-get-access-to-data/

NWEH

https://nweh.co.uk/

NorthWest eHealth was established in 2008 as a collaboration between Salford Clinical Commissioning Group, The University of Manchester, and Salford Royal Foundation Trust and provided the technology that enabled the Salford Lung study. NorthWest eHealth uses routinely-collected healthcare data for clinical trials.

The clinical trial platform enables feasibility, economic modelling, recruitment, real-time safety monitoring and data analytics to support the whole clinical trial lifecycle.

OpenSAFELY

https://www.opensafely.org

https://docs.opensafely.org/getting-started/

OpenSAFELY is a secure, transparent, open-source software platform for analysis of electronic health records data. It is currently running on pseudonymised primary care data from 58 million people's health records from practices using TPP and EMIS software. This data is accessed for the purpose of identifying medical conditions and medications that affect the risk or impact of COVID-19 infection on individuals, and identifying risk factors associated with poor patient outcomes as well as information to monitor and predict demand on health services. There are no industry research projects approved.

All platform activity is publicly logged. All software for data management and analysis is shared, automatically and openly, for scientific review and efficient re-use.

RCGP - Oxford/Nuffield ORCHID

https://orchid.phc.ox.ac.uk/index.php/orchid-data/

The ORCHID hub contains data aggregated from 1,800 general practices in England covering approximately 8 million patients who are broadly representative of the English general population. The Oxford-RCGP RSC team transforms routine clinical data from individual patient records at practice level into an accessible repository of data for health research. Data is transformed to a series of themed datasets, readily available for research.

ORCHID data is available for researchers interested in conducting primary care or linked data studies. Researchers should fill in the data request form (https://salesforce-eu.123formbuilder.com/form-42306/collaborator-request) and email to orchid-reg@phc.ox.ac.uk.

SeRP/NI Honest Broker Service

https://serp.ac.uk/
https://hscbusiness.hscni.net/services/2454.htm

The Honest Broker Service Service is the Trusted Research Environment for Health and Social Care (HSC) Northern Ireland. The HBS provides approved researchers with access to linked, de-identified health data via a safe setting. The HBS works under the Five Safes Framework to protect data confidentiality. All research projects are approved by the Honest Broker Service Service Governance Board. Projects must relate to Health and Social Care and be in the public interest, in that they will support development/delivery of public policy.

Researchers carry out analysis on record level, linked data which has gone through a de-identification process whereby all real-world identifiers have been removed. Researchers can only access the data via the safe setting and the only information that can leave the safe setting is aggregate statistical data.

UK Biobank

https://www.ukbiobank.ac.uk/

https://www.ukbiobank.ac.uk/explore-your-participation/basis-of-your-participation

UK Biobank is a well-known and well-documented large-scale biomedical database and research resource, containing in-depth genetic and health information from 500,000 UK participants. The database includes Whole Genome Sequencing data on 200,000 UK participants, is regularly augmented with routine and ad-hoc data and is globally accessible to approved academic and industry researchers.

References

1. https://www.abpi.org.uk/publications/unlocking-the-promise-of-uk-health-data/
2. https://understandingpatientdata.org.uk/what-do-people-think-about-third-parties-using-nhs-data
3. https://www.healthdatagateway.org/
4. https://www.HDR UK.ac.uk/helping-with-health-data/health-data-research-hubs/
5. https://www.hra.nhs.uk/about-us/committees-and-services/confidentiality-advisory-group/
6. https://zenodo.org/record/5767586#.YbM2tvF_pqs
7. https://www.gov.uk/government/publications/accessing-public-health-england-data/about-the-phe-odr-and-accessing-data
8. **DOI:** 10.1177/2047487320942338

abpi
Bringing medicines to *life*

The Association of the British Pharmaceutical Industry

A company limited by guarantee registered in England & Wales number 09826787

Registered office 7th Floor, Southside, 105 Victoria Street, London SW1E 6QT

RMI-0154-0322

Design by tothepoint