

Understanding Patient Data Secure Data Environment Survey

On 24th May 2024, we wrote out to a group of the key Secure Data Environments that hold health data to gather information on the topics below to support with a new UPD webpage.

Name of SDE/TRE	Organisation responsible	Short description of the data it holds	List of outputs	List of approved projects (e.g., in progress or starting, therefore no outputs yet)	Funding*	Further information
Clinical Practice Research Datalink (CPRD) Trusted Research Environment	Medicines and Healthcare products Regulatory Agency (MHRA), an executive agency of the Department of Health and Social Care	A representative sample of pseudonymised primary care data from contributing UK general practices with almost 30% population coverage. Linked healthcare datasets available including secondary care and mortality (death) data.	<i>CPRD has only recently started onboarding projects to the TRE so there are no outputs yet. Outputs (peer-reviewed publications) will be published on the CPRD bibliography web page: Bibliography CPRD. This will not be specific to the TRE.</i>	Summaries of approved projects available on CPRD webpage: Approved studies using CPRD Data CPRD . This does not specify which are taking place within the TRE.	Estimated spend on the TRE so far is £4.5m. CPRD operates on a full cost-recovery model via data access licence fees (this includes recuperating the costs of the TRE). Prices for researchers to access CPRD are here: https://www.cprd.com/pricing	CPRD Trusted Research Environment CPRD
SAIL Databank	Swansea University	<p>SAIL Databank contains anonymised (at the point of access to the researcher) health and administrative data for the population of Wales, some dating back roughly 30 years.</p> <ul style="list-style-type: none"> • 100% of secondary healthcare • 85% of primary healthcare • Disease and condition-specific datasets • Social care • Census data • Employment • Housing • Justice 	A list of publications can be found at: https://saildatabank.com/data/publications/	A list of approved projects can be found at: https://saildatabank.com/data/data-use-register/	<p>SAIL Databank receives funding from the Welsh Government's Health and Care Research Wales and UK Research and Innovation's Economic and Social Research Council (ESRC). Both funding sources support the infrastructure costs (technical and administrative / governance) of SAIL Databank. ESRC funding provides estate costs to Swansea University, where SAIL is based.</p> <p>SAIL has been core-funded by Health and Care Research Wales since 2006. Income breakdown:</p>	https://saildatabank.com/

					<p>2006 – 2010: £1,053,312 2010 – 2013: £1,748,885 2013 – 2015: £1,335,432 2015 – 2018: £2,150,157 2018 – 2020: £1,373,112 2020 – 2025: £4,545,557.13</p> <p>Total: £12,206,455</p>	
National Genomic Research Library (NGRL)	Genomics England	<p>All the data held is de-identified, consented data which participants have agreed to provide.</p> <p>It includes information such as:</p> <ul style="list-style-type: none"> • data collected from genomic sequencing tests • diagnoses and medical test results from health records, such as blood test results • medical notes and relevant information from local or national disease registries • relevant imaging data from health records, such as MRI scans, X-rays or pathology images 	<p>Information on publications arising from research on National Genomic Research Library data can be found on Genomics England's research publications page: https://www.genomicsengland.co.uk/research/publications</p>	<p>Information on the status and details of registered projects is available to Discovery Forum and Research Network members via the research registry: Browse the Research Registry - Customer Self-Service (genomicsengland.co.uk)</p>	<p>Details on Genomics England's sources of funding and how it is allocated can be found in their annual report under 'Source and application of funding': Genomics England Annual Report.</p> <p>It isn't possible to split out funding for the secure data environment specifically.</p>	<p>https://www.genomicsengland.co.uk/</p> <p>https://www.genomicsengland.co.uk/patients-participants/data</p>
Honest Broker Service Secure Data Environment	Health and Social Care Business Services Organisation (BSO) Northern Ireland	<p>De-identified health and social care data for the population of Northern Ireland. Data includes:</p> <ul style="list-style-type: none"> • inpatient and outpatient (emergency department, maternity services, labs) • the integrated health and social care system • dental and ophthalmic (eye) treatments • mortality (death) 	<p>A list of published papers using the service can be found by clicking the last link on this page</p>	<p>The data use register can be found here</p>	<p>The Honest Broker Service is funded via the Northern Ireland Trusted Research Environment (NITRE), part of Health and Social Care Data Institute (HSCDI).</p>	<p>https://bso.hscni.net/directorates/digital/honest-broker-service/</p>

National Safe Haven (for Scotland)	Public Health Scotland	It holds approved data extracted from national health, social care and administrative datasets (in addition to researcher provided cohorts – data from people that have taken part in research). Data will include, for example, pseudonymised inpatient, outpatient, cancer, census and data from many other datasets.	eDRIS (electronic Data Research and Innovation Service) has little information on publication and whilst researchers are encouraged to reference the National Safe Haven in their work, many do not let them know of their published works once information (their charts, reports, etc) has been released from the National Safe Haven.	<p>Projects will go through the Health and Social Care Public Benefit and Privacy Panel (HSC PBPP) if they use national health data and the Stats Public Benefit and Privacy Panel if they use administrative data such as data from the Scottish Government or Census data. Where both types of data are needed, applications will need to go to both panels. For HSC PBPP, a list of approved projects can be found here: https://www.informationgovernance.scot.nhs.uk/pbpphsc/application-outcomes/</p> <p>The eDRIS reference is on the left hand side of each document and not all approved projects will use the National Safe Haven. For SPBPP the approved projects list can be found here, and again the eDRIS reference is on the left of each document and not all approved projects will use the National Safe Haven: Scottish Government statistics: request our data - gov.scot (www.gov.scot)</p>	The National Safe Haven funding is not singular and should include eDRIS and EPCC income for the running of the service. There is no single figure for the total cost available for release.	National Safe Haven (NSH) - Electronic Data Research and Innovation Service (eDRIS) - Data research and innovation services - Public Health Scotland
UK Biobank Research Analysis Platform	UK Biobank	UK Biobank contains the de-identified genetic, lifestyle, imaging and health information and biological samples of 500,000 UK participants. The participants were recruited between 2006 and 2010, aged between 40 and 69 years, from England, Wales and Scotland.	Researchers using UK Biobank data must put their findings in the public domain. By the end of 2023, there were over 10,000 publications based on UK Biobank data. Some publications are highlighted on the website , and all publications using UK Biobank data can be found by	There are currently around 2,800 active projects underway involving over 18,000 approved researchers from more than 50 countries. A full list of approved research projects can be found on the website .	<p>Estimated funding specifically for the Research Analysis Platform is £20m from Wellcome.</p> <p>More detailed information on funding is on the website.</p>	<p>About Our Data https://www.ukbiobank.ac.uk/enable-your-research/about-our-data</p> <p>The UK Biobank Research Analysis Platform: https://www.ukbiobank.ac.uk/enable-your-research/about-our-data</p>

			searching databases such as PubMed .			nk.ac.uk/enable-your-research/research-analysis-platform Basis Of Your Participation https://www.ukbiobank.ac.uk/explore-your-participation/basis-of-your-participation
Our Future Health Trusted Research Environment	Our Future Health	<p>There are four datasets available:</p> <ul style="list-style-type: none"> • participant data (registration, consent and demographic data) • the baseline health questionnaire data (self-reported health information, and data about their household, socioeconomic status, work, education, and family history) • the genotype array data (genetic information from blood tests) • linked health records data (hospital admissions, outpatient appointments, accident and emergency attendances) and death data from the Office for National Statistics <p>Full information can be found here, and information on the data releases (updates every quarter) can be found here</p>	These will be signposted from https://research.ourfuturehealth.org.uk when available.	There are three ongoing studies which can be found here .	Our Future Health in general has received £79m from UK Research and Innovation (UKRI) initially, followed by another £51m from them in a later government budget. Separate to this they have raised £180m life sciences companies and health research charities. Information on funding is available on the website , but a breakdown is not available into the costs of the Trusted Research Environment versus other activities such as participant recruitment.	https://research.ourfuturehealth.org.uk/ https://ourfuturehealth.org.uk/protecting-your-data/how-we-make-data-available-for-research/

OpenSAFELY (the NHS England OpenSAFELY Service in OPENSafELY-TPP and OpenSAFELY-EMIS)	NHS England is the data controller for the platform and service; GP practices remain data controller for the electronic health records data; GP system providers (TPP and EMIS) are data processors; and the Bennett Institute at University of Oxford are also data processors and lead on platform design, software, and service delivery.	<p>The tools sit across the primary care records data of nearly all current patients in England.</p> <p>Other data, linked to this, includes: A&E, inpatients, outpatients, hospital administrative data, mortality (death), Covid tests and vaccinations, and others:</p> <p>https://docs.opensafely.org/data-sources/</p>	<p>As of June 2024, the OpenSAFELY user community across 22 organisations (e.g. universities and government bodies) has produced:</p> <ul style="list-style-type: none"> 74 academic publications: https://www.opensafely.org/research/ 20 pre-prints (awaiting publication): https://www.opensafely.org/research/ 23 published reports (with graphs, dashboards and tables): https://reports.opensafely.org/ 323 repositories of study code (analytical code used to do analyses): https://github.com/opensafely 	<p>Currently 155 projects approved by NHS England from 22 different organisations: https://www.opensafely.org/approved-projects/</p>	<p>Estimated total spend on OpenSAFELY platform since 2020 to date: £8-9m.</p> <p>This is made up of:</p> <ul style="list-style-type: none"> Wellcome - £2,277,389 NHS England - £2m National Institute for Health and Care Research / Medical Research Council together – £600k NIHR independently - £550k MRC independently - £2.56m <p>More detail was provided on the specific grants and estimates of how much per grant was spent on the TRE.</p>	<p>https://www.opensafely.org/</p> <p>https://www.youtube.com/watch?v=GRjRqOAlVy8</p>
NHS England Secure Data Environment Network	NHS England is the funder of the SDEs and oversees the programme. NHSE is entirely responsible for the NHSE Secure Data	<p>The National SDE holds data such as Covid-19 data, Hospital Episode Statistics, Maternity Services Data Set, Medicines Dispensed in Primary Care, and others. A full list is here.</p> <p>The regional ones vary and contain a range of specialist and local datasets, depending on local agreements. Each SDE will publish</p>	<p>Outputs from the National SDE are included in NHS England's data uses register: Data Uses Registers - NHS England Digital. This does not specify which data sharing requests have been met in the SDE specifically.</p> <p>The regional SDEs are in varying stages of</p>	<p>The NHS England SDE publishes approved studies through its Data Uses Registers - NHS England Digital</p> <p>The regional SDEs are in varying stages of development but all will have published data use registers by the end of this financial year, which will be accessible through their own websites or, for all</p>	<p>The Network is largely funded through NHS England funding, but in some cases builds on pre-existing infrastructure and development.</p> <p>The Programme was funded with £175m in total, which funds the SDE Network, the Digitals</p>	<p>Research powered by data - NHS England Digital</p> <p>East of England Sub-National Secure Data Environment (eoe-securedataenvironment.nhs.uk)</p>

	<p>Environment. Regional SDEs hold responsibility locally through NHS organisations.</p> <p>There are 11 regional SDEs which are:</p> <ul style="list-style-type: none"> • East of England • East Midlands • Great Western • Kent, Medway and Sussex • London • North East and North Cumbria • North west • Thames Valley and Surrey • Wessex • West Midlands • Yorkshire and the Humber 	<p>information about the data it holds by the end of this financial year (2024-25).</p> <p>The data is:</p> <ul style="list-style-type: none"> • pseudonymised, so that any identifying (or personal) patient information such as names, addresses or NHS numbers are replaced with artificial, or 'pseudo', information • minimised, so that only relevant patient information is provided for research projects. 	<p>development but all will have published data use registers by the end of this financial year, which will be accessible through their own websites or, for all registers, through the NHSE website.</p> <p>Examples of outputs include</p> <ul style="list-style-type: none"> • Prediction 'calculators' to understand patients' risk of and reduce hospital re-admission following treatment for heart failure in the East of England SDE • Assessments of the effectiveness of Sotrovimab (a monoclonal antibody) when given with antiviral treatments for COVID-19 in a clinical trial in the London SDE 	<p>registers, through the NHSE website.</p>	<p>Service and other programme activity.</p> <p>NHS England » Investing in the future of health research: secure, accessible and life saving</p>	<p>Homepage - West Midlands Secure Data Environment (WMSDE) (westmidlandssde.nhs.uk)</p> <p>Developing the Great Western Secure Data Environment (SDE) - ARC West (nihr.ac.uk)</p> <p>Kent, Medway and Sussex Secure Data Environment for research - Sussex Health & Care (ics.nhs.uk)</p> <p>Funding awarded to London organisations to improve use of data for research – London Health Data Strategy</p> <p>Secure Data Environment North East and North Cumbria NHS (northeastnorthcumbria.nhs.uk)</p> <p>https://thamesvalleyandsurreyhealthandcaredata.nhs.uk/using-patient-data/thames-valley-</p>
--	---	--	---	---	--	---

						and-surrey-secure-data-environment/ https://wessexhealthpartners.org.uk/our-priorities/vision-mission-and-strategic-workstreams https://westmidlands.sde.nhs.uk/ https://arc-yh.nihr.ac.uk/research/projects/yorkshire-and-humber-secure-data-environment/
--	--	--	--	--	--	--

We also approached the Office for National Statistics, as the [Secure Research Service](#) and the [Integrated Data Service](#) are Secure Data Environments that hold some health data. The Secure Research Service gives accredited researchers secure access to de-identified, unpublished data in order to work on research projects for the public good. The SRS has been providing secure access to data for 20 years, and the IDS is the replacement service. The [IDS](#) has some health data, such births and deaths, and datasets relating to Covid (e.g. the Covid-19 Schools Infection Survey linked with Test and Trace data, deaths by Covid-19 vaccination status, Virus Watch). The Office for National Statistics has stated that it's no longer relevant for them to share details about the SRS, and going forwards they would like to provide more information about the IDS but as the service is in its infancy they are unable to provide the requested information. We look forward to receiving this information in the future.

*Funding – when we asked for this information, we said “we know the funding question might be difficult to answer, particularly as sometimes funding is received for analytical projects within the SDE/TRE as well as contributing to the running of the environment itself. However, at UPD we think it's important to be transparent about the costs that are associated with managing these environments (which are likely to be more expensive than traditional data dissemination methods). We're only asking for estimates.” As we gave quite a broad definition, and different organisations took different approaches to answering this question, this has ended up being very broad estimates for only a few of the SDEs. This information should also not be read as a comparison between different SDEs; they operate on different scales, with different technologies, with different levels of staffing, and have different usages. We appreciate the efforts some organisations went to in order to provide detailed estimates of the money spent on their SDEs.