North West
SECURE DATA
ENVIRONMENT

NHS
Cheshire and Merseyside

Cheshire and Merseyside Secure Data Environment: FAQs

Version control:

Version	Date	Description
1.0	21 December 2023	Initial release

Policy and benefits

What is the Cheshire and Merseyside Secure Data Environment?

The Cheshire and Merseyside Secure Data Environment (SDE) is where NHS Cheshire and Merseyside securely stores data from GP practices and other providers to be used for insight-driven local service management.

The Cheshire and Merseyside SDE is the new name for the data store that has been in place since 2020 as part of the Combined Intelligence for Population Health Action (CIPHA) programme. The reason for this name change is because of how the Cheshire and Merseyside SDE will support the new North West Secure Data Environment (NWSDE) in the NHS Secure Data Environment Network.

A new data sharing agreement is being sought to cover this change, with the NWSDE providing secure access to de-identified data for research purposes.

The Cheshire and Merseyside SDE is also part of a new programme called ‘Data into Action’ that will co-ordinate the use of data across the Cheshire and Merseyside Integrated Care System (ICS) and with academic partners.

The Cheshire and Merseyside SDE, contracted by NHS Cheshire and Merseyside, is hosted by the Data Service for Commissioners Regional Office (DSCRO) and provided by Arden and Greater East Midlands Commissioning Support Unit (Arden & GEM). It is intended to serve NHS Cheshire and Merseyside, commissioner, local authority, provider, voluntary, community and social enterprise (VCSE), and research users.

The SDE is Cheshire and Merseyside’s recognised platform for conducting all forms of analysis, including population health management, risk stratification, planning, and evaluation and research.

The Cheshire and Merseyside SDE is managed by the NHS and it builds on existing research partnerships between health, local authorities and universities. It is supported by academic institutions within the region, including the University of Liverpool and Edge Hill University.

The aim of the Cheshire and Merseyside SDE is to offer near-real time, privacy-protecting access to rich linked data spanning different types of data captured for the purpose of:

- epidemiology reporting
- predicting outcomes and population stratification of vulnerable populations
- planning services and understanding future service provision
- evaluating and understanding causality
- research into novel interventions.

The Cheshire and Merseyside SDE covers a connected population of 2.7 million. The region includes 375 NHS partner organisations (including GP practices and NHS trusts) and nine local authorities situated in urban and rural environments.

The lead organisation for the Cheshire and Merseyside SDE is NHS Cheshire and Merseyside, which will work in collaboration with existing organisations strictly under the condition of the Data Sharing Agreement. These organisations include Health Innovation North West Coast (Academic Health Science Network for the North-West Coast); Clinical Research Network North-West Coast (CRN NWC); National Institute for Health and Care Research Applied Research Collaboration - North-West Coast (NIHR ARC NWC); and Health Data Research UK (HDR UK) North.

What is the North West Secure Data Environment (NWSDE)?

The Cheshire and Merseyside SDE feeds into the NWSDE, which includes data from:

- Cheshire and Merseyside
- Greater Manchester
- Lancashire and South Cumbria.

The NWSDE is one of 11 sub-national SDEs in the national NHS Secure Data Environment Network. The NWSDE covers over 7 million people, whilst preserving connectivity to local communities and clinical teams. Working at scale it will give researchers access to a much richer data environment, with the depth and breadth of data to feed innovation and research.

The data made available to NWSDE is the same data managed at Cheshire and Merseyside level. It is processed and accessed in the same way. Access to the data is governed by the parameters outlined in the Data Sharing Agreement, which includes approval at the Cheshire and Merseyside level Data Asset and Access Group (DAAG). Please see further down in this document for more details on how the data is processed and how access is governed.

What is the national Secure Data Environment programme?

The national Data for Research & Development (R&D) programme is investing up to £200m in data infrastructure to improve secure access to data for research purposes. The programme is making England-wide investments in Secure Data Environments and has established the NHS Research Secure Data Environment Network. The programme vision is that, by March 2025, there will be a world leading, NHS-wide health data research infrastructure that enhances patient care, sustains the NHS, supports innovation and is understood and well supported by the public. More information on the National Programme can be found on the NHS England - Transformation Directorate’s website.

What national policy drives the Secure Data Environment investment?

The Goldacre Review

Professor Ben Goldacre was commissioned by the Government in February 2021 to review how to improve safety and security in the use of health data for research and analysis. The report makes recommendations to benefit patients and the healthcare sector. More information and the full Goldacre Review is on the Government’s website.

Data Saves Lives strategy

The Data Saves Lives Strategy, published in June 2022, informs the Data for R&D programme. It outlines how data will be used to benefit all parts of health and social care. The strategy commits to investing in improving the quality and accessibility of at-scale data assets that make the most of health and care data, and to act on the findings and recommendations of the Goldacre Review, published in April 2022. It includes making a strategic move away from a system of data sharing for research and analysis using NHS data to one of controlled data access.

How is the Federated Data Platform (FDP) different?

The Federated Data Platform (FDP) is not a national data collection. It is software that will help connect disparate sets of data and allow them to be used more effectively for care. It will sit across NHS trusts and integrated care systems (ICSs) allowing them to connect data they already hold in a secure and safe environment. GP data will not be part of the national platform.

The software will be ‘federated’ across the NHS. This means every NHS trust and integrated care board (ICB) will have its own version of the platform that can connect and collaborate with other data platforms as a ‘federation’. This makes it easier for health and care organisations to work together, compare data, analyse it at different geographic, demographic and organisational levels, and share and spread new effective digital solutions. NHS England is not mandating the use of nationally procured software. NHS England will work with ICBs and NHS trusts to maximise opportunities to improve outcomes for patients.

On 21 November 2023, the contract to provide the software was awarded to a group led by Palantir Technologies UK, with support from Accenture, PwC, NECS and Carnall Farrar. There will be a six-month implementation period where products supported by the current platform will be transitioned across. Additionally, NHS England is expected to begin national public engagement in January 2024, ending in March 2025. Given those timescales, NHS Cheshire and Merseyside will consider how best to interact with the FDP and communicate this in due course. All decisions will be made in the best interest of patients, clinical effectiveness and patient safety.

For up to date information on the Federate Data Platform, visit the NHS England website.

How does the FDP relate to Cheshire and Merseyside SDE?

The way the FDP products are deployed is not uniform across every ICB. NHS Cheshire and Merseyside is already a great example of how leaders and providers of healthcare within integrated care systems (ICSs) have their own data platforms that are successfully supporting local leaders and clinicians to make decisions that improve patient care, staff experience and health outcomes.

NHS Cheshire and Merseyside already has a mature data management system, CIPHA, that is being developed into the Cheshire and Merseyside Secure Data Environment (SDE). This is separate from the FDP. It is owned and driven by the local data controllers, data processors, and patients and the public within Cheshire and Merseyside.

Cheshire and Merseyside SDE is enacting plans with its data controllers – NHS trusts, local authority providers and GPs – to engage them with the use of their asset. In parallel to that, a Cheshire and Merseyside patient engagement and awareness campaign has begun, including a number of public forums to ensure that the public understand how Cheshire and Merseyside uses data and how it securely manages its use.

What are the benefits for patients?

Research is a critical first step for identifying new ways to care for patients. Sharing and using data more effectively across Cheshire and Merseyside is already helping us to provide better services for our population's needs. Extending this to wider research, through ethical frameworks of publication and peer review, has the potential for more patients to benefit from other innovations sooner. Research and evidence gathered now will shape the NHS services that patients access in future.

What are the benefits for service providers, including GPs?

It is hoped that Cheshire and Merseyside SDE research projects will provide learning that can be integrated into routine care and that can improve patient outcomes as well as provide a platform and support for GP practices, primary care networks and community providers to be more involved in research.

Are there examples locally of where data has been used in this way to benefit patients?

During the COVID-19 pandemic and subsequent recovery, some COVID-19 related research has already been undertaken. This was conducted under the governance of the Control of Patient Information (COPI) Notice for COVID-19. Examples of these research use cases include:

1) SMART Testing

On 3 December 2020, "community testing" for SARS-CoV-2 antigen among people without symptoms of COVID-19 was expanded from Liverpool City to the wider Liverpool City Region (LCR). Between 3 December 2020 and 31 July 2021, 668,243 (45%) LCR residents aged five years and older had a registered result from a SARS-CoV-2 rapid antigen lateral flow test (LFT) performed at a testing centre or via a universal access home test kit. The Cheshire and Merseyside data was used to analyse this programme. Details of this research are described in this report on the Government's website.

2) Events Research Programme (ERP)

The ERP was a study that pioneered the return of events in a structured, scientifically and ethically robust manner to enable events to return at a scale not previously trialled. The objectives of the ERP were to build evidence on the risks associated with events-related transmission routes of the COVID-19 virus; the public health characteristics of events and surrounding activities; and the extent to which risk-mitigation measures can be implemented. Between 17 April and 15 May 2021, Phase 1 of the ERP conducted nine pilots, some running across multiple days, in a variety of indoor and outdoor settings, with variations of seated, standing, structured and unstructured audience styles, and a range of participant numbers. Cheshire and Merseyside data underpinned the research in the events in Liverpool, including a concert in Sefton Park, nightclub events and a business event. More information on this research is on the Government's website.

These projects demonstrate not only the research value of this data, but also how that research can positively benefit our population.

Population Health Management

Data has also been used extensively for the purposes of Population Health Management where data controllers (GP practices and NHS trusts) have signed the CIPHA Population Health Data Sharing Agreement. Various intelligence reports are available in the CIPHA system that allow clinicians and health and care workers to identify people who would benefit from direct care interventions, including in the following areas.

1) Fuel poverty

The Fuel Poverty dashboard identifies individuals who are at risk of poor health outcomes due to cold homes and fuel poverty. It is currently deployed into primary and community care teams so that treatment can be optimised, and warm home initiatives can be deployed. As an example, the St Helens Warm Homes for Lungs project has used this dashboard to identify and support 85 patients since launching in February 2023. These patients have been referred for wellbeing and warm home support, with 14 also being seen by the Pulmonary Rehabilitation team and 18 joining the COPD Telehealth Service. A total of 65 patients have received £500 payments from household support funds, with further payments being made in autumn/winter 2023. Similar services and outcomes are being achieved across NHS Cheshire and Merseyside.

2) Waiting list stratification

The waiting list stratification tool links primary, community and mental health data with waiting list data, to enable stratification of the waiting list population for risk of adverse outcome, deprivation and other protected risk factors. It is available to support prioritisation of waiting lists and also inform care providers which patients may benefit from a waiting well initiative.

3) Enhanced case finding

A tool to support the stratification of data is applied to identify and proactively support vulnerable patients who need multi-disciplinary teams to co-ordinate services around them. The tool being used is the Johns Hopkins Adjusted Clinical Groups (ACGs) system, which is widely used globally with many published papers on its effectiveness for risk stratifying health data sets. (Example: Association between Johns Hopkins Adjusted Clinical Groups risk scores and self-reported outcome measures: an observational study among individuals with complex or long-term conditions in Norway | BMJ Open.)

Technical, data and information governance

What, technically, is the Cheshire and Merseyside SDE?

The SDE is an Azure Cloud Data Management Environment. It contains several software components including a cloud warehousing solution, with various analytical and querying tools and also a Power BI front end to visualise data. The technical environment provisions data for projects on a project-by-project basis, ensuring data is minimised for the specific purpose. Data is provisioned in an ‘airlock’ system meaning any analysis needs to take place within that environment and no data can leave the environment without approval and meeting certain minimisation criteria.

What data is held in the Cheshire and Merseyside SDE?

Building upon recognised routes for the processing of NHS England data into Arden and GEM, the SDE combines standard national commissioning datasets with additional local data flows.

Data will flow from three core data sources:

1. National Commissioning Datasets from NHS England.
2. Local organisational data flows direct from NHS providers of services.
3. Data from the Cheshire and Merseyside Shared Care Record provided by Graphnet as a data processor.

NOTE: Data already flows in this way for the purposes of Population Health. The new Cheshire and Merseyside SDE Data Sharing Agreement will allow this data to be used for research in academia.

Can users see identifiable data in the SDE?

No. All data in the SDE is de-identified, which means that names and addresses are removed. Date of birth is changed to age and the post code is shortened to provide high-level geographical analysis only. Within the project work spaces, patients cannot be identified.

Who will have access to the Cheshire and Merseyside SDE?

There are two user groups that will have access to data provided by the Cheshire and Merseyside SDE:

1. NHS Cheshire and Merseyside analytical teams, including those with honorary contracts.
2. Researchers from academia for approved research projects.

Will industry and charities be able to access the SDE?

Not at this stage. The Data Sharing Agreement SDE ‘Research in Academia’ only covers researchers from academia to access the SDE. A separate Data Sharing Agreement will be developed for access to the data by commercial partners such as pharmaceutical companies and technology companies. The development of this separate data sharing agreement will be done in conjunction with the public and data controllers.

Will the SDE charge for data access?

National funding is available until March 2025 to set up, develop and run the Secure Data Environment. During this time, we are working with national colleagues on a commercial model to charge for data access. This is a 'cost recovery' model. The money is intended to go towards the running of the SDE for future years, so it is self-sustaining. This is similar to other research data environments available, such as collections through the Health Data Research UK Research Gateway and Clinical Practice Research data link (CPRD)

How is access granted to the Cheshire and Merseyside SDE?

Data Access approval is on a project-by-project basis. Applicants must apply to the Cheshire and Merseyside Data Asset and Access Group (DAAG) that meets every three weeks and includes named members from each of the data controllers, each of the data processors, a data protection officer, Caldicott Guardian, academic partners and the public.

Applicants must complete a comprehensive Data Access Request Form, which has sections on project description, benefits, intended outputs, data specification, legal basis, information governance, and patient and public involvement.

The group ensures requests to use the stored data maintain the integrity and purpose of the specific Data Sharing Agreement. The group will ensure the appropriateness of the Role Based Access Control (RBAC) framework in terms of individuals and groups with access. The group's functions from the terms of reference are below:

- Oversight of the Data Access Request Process, approving data access requests from organisations.
- Ensuring all information governance requirements are met, including UK General Data Protection Regulation (GDPR), the Data Protection Act 2018,Common Law Duty of confidentiality, Caldicott principals, data minimisation and public benefit.
- Ensuring programmes applying have undertaken appropriate patient and public involvement and engagement in their design.
- Ensuring individuals and organisations applying have met the required conditions for data access.
- Ensuring technical specifications of data and technology are accurate and data is minimised to project specific requirements.
- Oversight of the development of the process, ensuring the process is developed in line with any changes to national policy or data sharing arrangements and escalating to the NHS Cheshire and Merseyside Information Governance Committee where changes need to happen.
- Ensuring applications have adequate scientific critique of research proposals.

What information governance controls are in place?

There are two information governance gateways that applicants need to achieve:

1) Organisational Information Governance Gateway

The employing organisation or department within an organisation of the individual researcher applying for the data will be required to demonstrate achievement of one of the following:

- Data Security Protection Toolkit, or
- Cyber Essentials Plus, or
- Equivalent ISO standard.

2) Individual Information Governance Gateway

The individual will be required to undertake the Safer Research Training and sign an individual contract detailing the parameters under which the data is being used: CIPHA SDE Terms & Conditions of Access Agreement. See Become an accredited researcher - Office for National Statistics (ons.gov.uk)

How is this compliant with the General Data Protection Regulation?

Below is the lawful basis for processing under UK GDPR.

Processing Personal Data - Article 6

6(1)(e) processing is necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller.

Processing Sensitive Personal Data – Article 9

9(2)(h) processing is necessary for the purposes of preventive or occupational medicine, for the assessment of the working capacity of the employee, medical diagnosis, the provision of health or social care or treatment or the management of health or social care systems and services on the basis of domestic law or pursuant to contract with a health professional and subject to the conditions and safeguards referred to in paragraph 3.

9(2)(j) processing is necessary for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes in accordance with Article 89(1) (as supplemented by section 19 of the 2018 Act) based on domestic law which shall be proportionate to the aim pursued, respect the essence of the right to data protection and provide for suitable and specific measures to safeguard the fundamental rights and the interests of the data subject.

How is the Common Law Duty of Confidentiality satisfied?

For research, the Common Law Duty of Confidentiality requires that there should be no use or disclosure of any confidential patient information for any purpose other than the direct clinical care of the patient to whom it relates unless:

- The patient explicitly consents to the use or disclosure.
- The disclosure is required by law.
- The disclosure is permitted under a statutory process that sets aside the duty of confidentiality.

For local flows coming from the Graphnet/System C supplied Shared Care Record, data is de-identified under a Confidentiality Advisory Group (CAG) approved Section 251 for using data for research: Confidentiality Advisory Group - Health Research Authority (hra.nhs.uk)

National data is de-identified within Data Service for Commissioners Regional Office (DSCRO), which operates under an NHS Direction from the Secretary of State Secretary of State Directions (Secretary of State Directions - NHS Digital).

Data is appropriately de-identified and, therefore, at the point of access by planners and researchers, is not owed a duty of confidentiality.

Is there a Data Protection Impact Assessment?

Yes, a full Data Protection Impact Assessment has been completed and can be accessed on the NHS Cheshire and Merseyside website.

What is a Tier Two Data Sharing Agreement?

There are three tiers to the Data Sharing Agreement framework across Cheshire and Merseyside.

- **Tier Zero: memorandum of understanding**. This is an agreement in principle to share information with partner organisations in a responsible way. This is signed by a chief executive (or equivalent) and is required regardless of further agreements.
- **Tier One: Data Sharing Agreement – standards**. This outlines agreed procedures for sharing confidential information, setting the minimum requirements for all participating organisations. This is signed by the designated responsible officer in each partner organisation.
- **Tier Two: Data Sharing Agreement**. This is a template for the safe sharing of personal data. The agreement shows what information should be shared and how, under what circumstances by whom, and is tailored to individual partnerships/programmes/projects. Tier Two agreements are signed by the Senior Information Risk Owner (SIRO) and/or Caldicott Guardian, or alternatively, chief executives (or equivalent) at partner organisations.

What Tier Two Data Sharing Agreements are in place across Cheshire and Merseyside?

There are currently several Tier Two Data Sharing Agreements that have been approved for Cheshire and Merseyside, which include:

- Tier Two for Direct Care
- Tier Two for Population Health
- Tier Two for Research in Academia (approved but data controller sign-up pending)

Who are the data controllers?

The data controllers are:

- NHS Cheshire and Merseyside
- Acute providers within NHS Cheshire and Merseyside
- The local authorities within Cheshire and Merseyside;
- North West Ambulance Service (NWAS)
- GP practices within NHS Cheshire and Merseyside
- GP out of hours services
- The approved research organisations and researchers accessing the data (see above answer to *'How is access granted to the Cheshire and Merseyside SDE?'* for data access approval process)

Who are the data processors and where is data stored?

Data is processed and stored by:

- Graphnet System C who process data for secondary uses via the shared care record;
- Arden and GEM, who process national and local data for secondary uses via the Data Service for Commissioners Regional Office

Data is stored in the UK.

How is the data that leaves the SDE minimised and is there guidance for publishing?

The data used for research is only accessible in the following ways:

- All data is de-identified, so no SDE user could re-identify any individual.
- No individual person's data ever leaves the SDE.
- Only anonymised data, such as diagrams, charts or aggregated tables, is extracted from the SDE for use in research papers or outputs.

Statistical disclosure checks are required to ensure that any data, such as charts, diagrams or tables, are safe to export. This process will be audited and will include anonymisation techniques as suggested by the UK Data Service (Anonymising quantitative data — UK Data Service Anonymising qualitative data — UK Data Service).

This includes a list of primary anonymisation techniques as follows:

- Remove direct identifiers.
- Aggregate or reduce the precision of a variable.
- Generalise the meaning.
- Restrict the upper or lower ranges.
- Anonymise relational data.
- Anonymise geo-referenced data.

Once this aggregated data has egressed from the SDE it will be used for publication in academic papers, and to support other uses, such as determining health and care policy decisions, or the efficacy of new interventions.

Patients and public involvement and engagement (PPIE)

How will patients be informed about Cheshire and Merseyside Secure data Environment?

There will be a full public involvement and engagement campaign to inform people of the benefits of SDEs and the different opt out options available to them. Local events will be held to explain this work with resources, such as leaflets, web pages and social media content, being shared with patients and the public. Alongside this, regional and national engagement campaigns will also be explaining what SDEs are and how they will work. Those regional and national resources will also be available to support patients locally.

If you would like more information, please visit the NHS Cheshire and Merseyside website.

An Equality Impact Assessment is underway for the PPIE work which will ensure that the PPIE is representative of the groups in the population who are vulnerable and most at risk of inequality.

Can patients opt out of their data being in the Cheshire and Merseyside SDE?

Yes. Patients can opt out in the following ways:

- **GDPR right to object** – prevents data sharing for any reason, including between services providing your care.
- **Type 1 Opt-out** – prevents your GP practice sharing your data for anything except your care, except when it is required by law.
- **National data opt-out** – prevents your personal and healthcare information being used for research and planning. This can be viewed or changed at any time.
- **Cheshire and Merseyside local data opt-out** – allows you to opt out of individual studies.

A separate leaflet on the NHS Cheshire and Merseyside website explains these options in more detail.

Do we need a new privacy notice to inform patients?

Each data sharing partner will already have a privacy notice that explains to patients how their data is used. Partners should update their notice to make reference to the sharing of the data in de-identified form for the purposes of research, risk stratification and planning, noting the organisations that data is being shared with. Suggested wording can be found below:

Locally across Cheshire and Merseyside, data is being shared securely with a data processor called Graphnet and with Arden and GEM for the purposes of protecting public health, providing healthcare services to the public, risk stratifying population to target direct care, and planning health care services. No data that identifies a person will be used for purposes other than direct care. Data is also being shared securely via these data processors with academia for the purpose of healthcare research only. All data to support research is de-identified and does not identify individual patients. If you have previously opted out of data sharing, your data will not be used.