University of Sunderland

Sharing learning from **the processes involved in linking children and young people's multi-sectoral data** in Sunderland and South Tyneside

NOVEMBER 2025

Dr Sarah Martin-Denham, Nathan Scott,
Dr Karen Horridge, Anne Pinney, Sam Seldon,
Michael Ball and Richard Burns

The True Colours Trust

NIHR | School for Public Health Research

Contents

Executive summary

> **This briefing paper shares learning from the processes involved in linking data for two data sharing and linkage projects in Sunderland and South Tyneside in the North East of England between 2020 and 2025, led by Dr Sarah Martin-Denham, Associate Professor of Care and Education, University of Sunderland (UoS). Both projects were service evaluations and began without any previous data linkage work across health, education and children's social care in the local areas.**

This paper will not share findings from the two projects, only the processes required to undertake 'data linkage'. Data linkage consists of combining data records that relate to the same person, but come from different sources, to create a new, enhanced data resource. The present authors endeavoured to link various administrative data sets in this fashion, and use the resulting data resource as the basis for two service evaluation projects. The projects referenced throughout the paper are:

Project one: Exploring if children in Sunderland identified in school census and social care administrative data as having learning disabilities/learning difficulties had appropriate paediatric clinical assessments completed and reported across agencies.

Project two: Exploring the needs, pathways, risk and protective factors of children in Sunderland and South Tyneside who have ever been suspended or permanently excluded from school.

While intended primarily for local audiences, this briefing paper raises issues requiring national policy attention. The benefits of data linkage are far-reaching, but to achieve them, extraordinary determination, tenacity and sustained commitment from all project partners was required, in order to find solutions to the many and substantial barriers in accessing and linking multi-sectoral data.

The principal motivation for pursuing data linkage opportunities in Sunderland (as opposed to alternative local areas) was that the project partners could attest to the high quality and consistency of paediatric data collection by paediatricians at all points of outpatient care in the area, including on the hospital site, outreach clinics in special schools, and home visits.

The paediatric data in South Tyneside and Sunderland NHS Foundation Trust reflect the multi-faceted needs of each child, including primary and secondary diagnoses (as many as are present), family-reported needs, need (or not) for round-the-clock care and any technology dependencies.

The project team and partners began the information governance processes for both processes in 2020 believing that linking data would be straightforward. Initially with funding from the Department of Housing Levelling up, Housing and Communities (DLUHC) and the UoS a project team was created. This included representation from UoS, Together for Children (TfC), Sunderland City Council (SCC), Sunderland Parent Carer Forum, NHS North of England Care Commissioning Support (NECS) and South Tyneside and Sunderland Foundation Trust (STSFT). In addition, academic, practice, policy makers and public partners supported the project.

It has taken five years to advance the projects, which highlights the complexities of linking data across agencies. The challenges posed by information governance (IG) processes persisted and considerably impacted the timeliness of both projects. This briefing paper summarises the lessons learned from these projects.

Most of the challenges to sharing and linking data stemmed from obtaining the proper permissions from data controllers and processors across and between organisations. We learned that there are different routes to securing data for linking, local and national. Initially, both projects followed a national route to receive data from NHS England, formerly NHS Digital. However, over time, it became clear that the national route required more approvals than were needed through local arrangements.

Due to time constraints, project one followed a local route and project two a national route. At the time of publication, the research team has successfully linked local area health, education and children's social care administrative data for project one, and have begun the service evaluation process.

Policy developments such as the advancement of the Children's Wellbeing and Schools Bill offer much-anticipated opportunities for data linkage using national flows of data. However, the timeline and practical implementation of these changes remain unclear. We hope this briefing paper helps to inform the creation of simpler, more consistent pathways for others to reap the benefits of linking and analysing data, ultimately improving outcomes for vulnerable children and their families.

The following recommendations are based on the learning from the projects:

Recommendations

Recommendation 1	Use a consistent, unique identifier, such as the NHS number, to support efficient and effective data sharing and linkage.
Recommendation 2	Establish a national mandate and funding to link children with Special Educational Needs and Disabilities (SEND)'s health, education and social care data, with clear expectations on data sharing between health and LA services through compatible data systems.
Recommendation 3	Develop a joined-up approach from central Government and the inspectorates, so local areas can be confident that their investment in developing data linkage systems will meet expectations.
Recommendation 4	Fund opportunities for LAs and health partners to share learning about how they have successfully linked health and LA data to improve outcomes for children.
Recommendation 5	Invest in developing best practice guidance and IG templates to make it easier for ICBs and LAs to develop data sharing protocols and processes.
Recommendation 6	Invest in a consistent and easy to use data capture interface to improve documentation in electronic medical records of the multifaceted needs of everyone attending a clinical encounter in the health service, regardless of setting. This should be underpinned by a harmonised set of agreed terms linked to SNOMED-CT codes so that reliable comparisons about all needs can be made over time and across settings. These data would be stored in secure data repositories in NHS organisations and be available to inform service design, produce quality service improvements facilitate research, and for secondary uses such as reporting to NHS mandated data sets, which currently include the Community Services Data Set, a key dataset for understanding trends in SEND and the multi-faceted needs of disabled children and their families.
Recommendation 7	Mandate consistent data capture and documentation of all needs at all points of care and services across agencies to ensure that quality data are available to link and analyse to improve outcomes for children and young people, with a transparent process of consent in place for documentation, storage and use of these data.
Recommendation 8	Clarify the definitions of the different terms used across services and agencies to describe learning disabilities and learning difficulties and the assessment pathways that underpin these.

Acronyms

A&E	Accident and Emergency
AGD	Advisory Group for Data
CYPHS	Children and Young People's Health Services Data Set
CiN	Children in Need
CSDS	Community Services Data Set
DARS	Data Access Request Service
DfE	Department for Education
DLUHC	Department for Levelling Up, Housing and Communities
DPBD	Data Protection by Design
DPIA	Data Protection Impact Assessment
DSA	Data sharing agreement
DSPT	Data Security and Protection Toolkit
ECHILD	Education and Child Health Insights from Linked Data
FSM	Free School Meals
GDPR	General Data Protection Regulation
HES	Hospital Episode Statistics
HRA	Health Research Authority
ICO	Information Commissioners Office
ICB	Integrated Care Board
IGARD	Independent Group Advising on the Release of Data
IG	Information Governance
IAPT	Improving Access to Physiological Therapies Data Set
LA	Local authority
MHSDS	Mental Health Services Data Set
MLD	Moderate Learning Difficulty
NECS	North of England Care System Support
PMLD	Profound and Multiple Learning Difficulty
SCC	Sunderland City Council
SENCO	Special Educational Needs Coordinator
SEND	Special Educational Needs and Disabilities
SLD	Severe Learning Difficulty
STSFT	South Tyneside and Sunderland NHS Foundation Trust
SUI	Single Unique Identifier
TfC	Together for Children
UoS	University of Sunderland

Glossary of terms

Axym	Axym is a cloud-based data access platform that provides a 'Single Source of the Truth' for multi-organisations and users. With approval it gives access to health and care data for analysis, evaluation and research in a safe and secure data environment, providing a private ring-fenced space to enable users to access and interrogate their data
Batch tracing	In the present project, batch tracing refers to identifying individuals in local authority (LA) datasets and tracing them to demographic details collected through healthcare (GP registration) to retrieve NHS numbers and add them to their local authority data
CiN census	The children in need census (CiN) is an annual statutory census for all local authorities. It collects data on children referred to LA social care services because their health or development is at risk
Community Services Data set	A patient-level, output based, secondary uses (re-purposed for non-direct care activities like planning, research, or commissioning) data set which aspires to deliver 'robust, comprehensive, nationally consistent and comparable person-centred information for people who are in contact with publicly funded Community Health Services'
Data Access Request Service (DARS)	An NHS England service that following approval enables access to health and social care data for organisations external to the NHS
Data controller	The natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data
Data processor	The natural or legal person, public authority, agency or other body which processes personal data on behalf of the controller
Data Protection by Design	The UK GDPR requires appropriate technical and organisational measures to implement the data protection principles effectively and safeguard individual rights. This is 'data protection by design and by default'
Data Sharing Agreement	Data sharing agreements set out the purpose of the data sharing, cover what happens to the data at each stage, set standards and help all the parties involved in sharing to be clear about their roles and responsibilities
Data Protection and Impact Assessment	A Data Protection Impact Assessment is a self-assessment to identify risks arising out of the processing of personal data and to minimise these risks as far and as early as possible
Data Security and Protection Toolkit	The Data Security and Protection Toolkit is an online self-assessment tool that allows organisations to measure their performance against the National Data Guardian's 10 data security standards
ECHILD	The ECHILD project is a research study run by University College London that joins together existing health, education and social care information for all children in England for the first time
Independent Group Advising on the Release of Data	The Independent Group Advising on the Release of Data independently assessed applications for health data; now the Advisory Group for Data (the same role as IGARD)

Acknowledgements

Without the enduring contributions and goodwill of the North East project partners, we would not have successfully linked local area health, education and children's social care administrative data. Special thanks to the following collaborators, without whom data linkage would not have been possible:

- Sam Seldon, Ruth Reah and Kylie Harrison (UoS)
- Richard Burns (SCC, formerly TfC)
- Kevin Brown, Stacy Hodkinson, Andy McCoy and Jacob Donaghue (TfC)
- Mitchell Bates, Amy Davison and David Wilkinson (South Tyneside Council)
- Bob Gaffney, Victoria Campbell-Molloy, Michael Ball and Michelle Gaff (NECS)
- Matt Thubron and Jim Carroll (STSFT)

Thank you to the academic, practice, policy makers and public partners provided support for the project, in particular:

- Dr Karen Horridge
- Anne Pinney
- Professor Lorna Fraser
- Professor Dougal Hargreaves
- Maureen Morris MBE
- Parents of children at Portland School, Ascent Trust, Sunderland

We would like to extend our thanks to the funders:

Project 1: The UoS, The True Colours Trust and the National Institute of Health Research, School of Public Health Research*

Project 2: The former DLUHC, Local Data Accelerator Fund

*This project is supported by the National Institute for Health and Care Research (NIHR) School for Public Health Research (SPHR) (SPHR-CYP-WP1-CCD). The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care.

1.0 Background

The following section, in brief, details what data linkage is (1.1), the known benefits of data linkage (1.2) and the policy context (1.3).

1.1. What is data linkage?

'Data linkage combines data from different sources that relate to the same person to create a new, enhanced data resource' (Harron, 2022, p. 1). Linked data enables agencies to identify trends in service access and evaluate these services accordingly (Downs et al., 2017), answering questions that require large sample sizes and are especially powerful for capturing populations that are hard to reach (Chiu et al., 2016; Aldridge et al., 2019). Through linked data, you can explore the relationships between education, social care, and healthcare outcomes (Fleming et al., 2021). Yet, routinely collected data by education, health and social care are rarely linked by local authorities (LAs) (Martin-Denham et al., 2023) or Integrated Care Boards (ICBs) (Scott et al., 2024).

In 2023, 73% of local authorities who responded to an FOI (N=91) were linking their own administrative datasets, but only 24% of these were also linking their data with healthcare data.

Similarly, 58% of ICBs who responded to an FOI (N=31) were linking their internal datasets, but only 16% were linking their data with LA data.

Both local authorities and ICBs primarily used data linkage for strategic planning, and the creation of data dashboards, while some also used it for joint commissioning, and service reviews.

Martin-Denham et al. (2023); Scott et al. (2024)

1.2. The known benefits of data linkage

The breadth and depth of local data sets provide opportunities to determine the impact of interventions to improve outcomes for children (ADR UK, 2023). Linking administrative datasets can help identify those at risk of adverse outcomes, enable early intervention, and gather information about specific groups (Atherton et al., 2015). Local authorities and other organisations can use this knowledge to improve children and young people's quality of life in a variety of ways, including producing data dashboards, strategic planning, targeted interventions and joint commissioning (Downs et al., 2017; Sohal et al., 2022; Martin-Denham et al., 2023; Scott et al., 2024; Pinney et al, 2024).

Data linkage also enhances the decision-making evidence base by drawing together a larger, richer dataset. During the COVID-19 pandemic, for example, the lack of ethnicity information on death registrations was identified and rectified by linking this data with the 2011 census (GAD, 2021). Linking data between healthcare and education records can provide useful ways of examining the impacts of mental health and illness on academic achievement (Downs et al., 2019).

Another indirect benefit of data linkage often observed is the improvement of administrative data quality across all involved services. Because data linkage projects are highly dependent on rich, high-quality data from the administrative datasets being linked, gaps, inconsistencies, errors and data blockages are often resolved through data checks and processing. Resolved errors include duplication, technical glitches, and missing data (Holman et al., 2008). This was the case in the data linkage projects informing this briefing paper. Specifically, the Community Services Data Set (CSDS) was found to be unpopulated, and a blockage in the flow to this data set was resolved.

There have been initiatives to link children's data nationally. The most noteworthy of these initiatives is University College London's Education and Child Health Insights from Linked Data (ECHILD), in partnership with DfE, NHS England and the Office for National Statistics (ONS). This study was the first to combine health, education and social care information for all children in England. If the ECHILD datasets are sufficient and timely, local areas may not need to undertake data linkage themselves, as they could extract their own local area's statistics from the ECHILD datasets. However, it currently does not include key data required for the Sunderland and South Tyneside projects, notably data about the multi-faceted needs of all children and young people seen in clinics across hospital and community settings. Since these projects were undertaken, the CSDS has been added to ECHILD. However, this would not include the diagnostic breadth and depth of data about individual, multi-faceted needs, as documentation of these relies on systems being available at all points of clinical care across all providers and settings to capture and record them easily. While these data should be collected and processed through the CSDS, it was found to be residing in an acute outpatient dataset in Sunderland and South Tyneside. This further highlights the barriers posed by inconsistencies in administrative data collection.

Another promising initiative is the Pan London Data Sharing Agreements Project, which was delivered by the London Office of Technology & Innovation with the London Safeguarding Children Partnership. The initiative has successfully signed all 32 London boroughs to a 'London Multi-Agency Safeguarding Data Sharing Agreement for Safeguarding and Promoting the Welfare of Children'. The agreement is yet to fully realise its possible benefits; however, the London borough responses suggest some early positive impacts, including:

- Consistency of approach and a reduction in multiple formal agreements that run the risk of creating confusion
- Builds good working relationships and gives confidence to all professionals
- Reliable framework
- Efficiency of process

In 2023, Pinney et al., funded by UKRI Policy Support Fund and the UoS, approached LAs and invited them to share outcomes of data linkage endeavours. This study generated four case studies:

1. In Bedford Borough, NHS data and Education Health and Care plans (EHCPs) were linked, resulting in the local authority identifying delays in accessing speech and language therapy, long waits between appointments and some children being incorrectly discharged from services.
2. Bedford Borough also used data linkage to identify children with Special Educational Needs and Disabilities (SEND) in health data, revealing that children were more likely to access core Child and Adolescent Mental Health Services (CAMHS), rather than the neurodevelopmental team within CAMHS. This led to changes within CAMHS, including more SEND staff training, and a lightening of paediatrician caseloads.
3. Hartlepool Borough Council has an integrated 0-19 service for health visiting, school nursing and Early Help, which uses NHS numbers as consistent identifiers across the two information systems used by these teams. A data sharing agreement is in place to allow data to be shared on babies identified with a disability, antenatally or after birth. This enables the local authority to ensure specialist support is in place when the time is right for the child to start at nursery or primary school.
4. Bradford Council has utilised linked data to create a live data dashboard that refreshes daily and incorporates various LA data sets. This dashboard enables the identification of vulnerable children and young people via their SEN support status, social care status, and presence within youth offending data. This data dashboard has allowed Bradford Council to 'monitor changes in vulnerable cohorts' and was crucial during the COVID-19 pandemic in keeping track of vulnerable children and young people.

1.3. Policy context

The SEND review (DfE, 2022) proposed to make better use of data in the SEND system as current data capture is 'inconsistent' and does not 'enable local systems and leadership to respond to local needs before it is too late' (p. 69). The SEND and Alternative Provision Improvement Plan (HM Government, 2023) set out proposed national and local inclusion dashboards to present performance data across education, health and care 'and provide a basis for measuring whether we are achieving our mission of improved outcomes, better experiences and a financially sustainable system'. The Independent Review of Social Care (MacAlister, 2022) recommended that the DfE should have a proactive strategy on making better use of data in children's social care, including a strategy for data linking with other data sources such as education, hospital and justice data. The Labour manifesto (2024) set out to improve data sharing across services, with a single unique identifier, to better support children and families.

Policy discussions around data linkage recognise that England's lack of a unique child identifier is one of the first hurdles to overcome. The parliamentary debate on the Health and Care Act (2022) prompted the Government to develop proposals to improve data sharing between health and social care to safeguard children and to explain their approach to implementing a consistent child identifier. Their report to Parliament (DfE, 2023) cautiously supported using the NHS number as a consistent child identifier, proposing regional pilots and further work to improve the 'interoperability' of data systems and to build practitioners' confidence in sharing information. This remains an active focus of policy development under the Labour government elected in July 2024, following commitments made in their health and education strategies in the run-up to the general election. For example:

> **'Labour will improve coordination between education, social care and the wider services that support families by piloting the expansion of a children's number like the NHS number that stays with children not just for their school career but for their whole childhoods, ensuring that their needs are better met, and any issues are addressed early'** (Labour Party, July 2024).

While in government, the Labour Party have begun the process of passing the 'Children's Wellbeing and Schools Bill', which includes details on a 'consistent identifier'. At the time of writing, this bill is currently at the committee stage in the House of Lords, but if the bill is passed, a legal requirement would be established that, for the custodians of certain types of data, 'the designated person must include the consistent identifier in the information processed'. This is caveated by the requirement that this consistent identifier can only be included in children's data when it 'is likely to facilitate the exercise by any person of a function of that person that relates to safeguarding or promoting the welfare of children' (Long, Roberts and Foster, 2025).

The Department for Education and local authorities are grappling with rising demand for high needs provision, putting pressure on local budgets (Parish et al., 2019). Furthermore, many studies have evidenced that those with SEND are exposed to increased risk of poverty, unemployment, poor housing and discrimination (Parsons and Platt, 2022; Krasniki, Carr and Stevens, 2023). Despite this, children's multi-faceted needs are often not accurately captured and described (Horridge, 2018) nor are they similarly visible as data are often collected in silos (Wilcock, Elliot and Symons, 2022).

2.0 The data linkage projects related to this briefing

The following section outlines the aims and objectives of project one (2.1.1), the datasets for project one (2.1.2). Section 2.2.1 details the aims and objectives for project two and 2.2.2 the datasets relating to the project.

Project one: Exploring if children in Sunderland identified in school census and children in need administrative data as having learning disabilities/learning difficulties had appropriate paediatric clinical assessments completed and reported across agencies.

Project two: Exploring the needs, pathways, risk and protective factors of children in Sunderland and South Tyneside ever suspended or permanently excluded from school.

2.1. Project one

This project investigates whether children in Sunderland identified in the school census and children in need administrative data as having learning disabilities/learning difficulties have had appropriate paediatric clinical assessments completed and reported across agencies.

2.1.1. Project one: aims and objectives

Aims

1. Develop a linked cross-agency dataset for children to determine if those identified as having learning difficulties/learning disabilities have had a clinical paediatric assessment and investigations
2. Produce a publicly available manual documenting the best practice in linking health, social care and education data
3. Form and collaborate with a national network to undertake the project and future research

Objectives

1. Collaborate with a national network to influence and provide a critical voice to the research
2. Amend data sharing agreements so specific health, education and social care datasets can be linked and used for research purposes
3. Produce individual-level linked health, education and social care datasets for all children living in Sunderland from 2018/19 to 2023/4
4. Determine the rate of children identified as having a learning disability in paediatric data who do not appear in school census and/or CiN census data
5. Determine the rate of children with a learning disability in the CiN census who appear in school census data as having a learning difficulty
6. Determine the characteristics of children with learning disabilities in the paediatric data compared to their learning difficulties recorded in the school census data
7. Assess whether any clinical features (e.g. severity of learning disability, other medical conditions) or demographic characteristics (e.g. school type, deprivation status) are associated with receiving a paediatric medical assessment
8. Make recommendations locally and nationally about future data collections in each sector to ensure children's needs are fairly/appropriately represented
9. Write a publicly available manual documenting the process and best practice to linking education, health and social care data

2.1.2. Project one: the datasets

Table 1 provides detail on the administrative datasets included for project one.

Table 1. Health, education and children's social care datasets: project one

Disabilities Terminology set matched to SNOMED CT	School Census	AP Census	SEN2	Children in Need Census
Age/DoB, gender, ethnicity (all datasets)				
Lower layer Super Output Areas of residence (small geographic zones used to publish local statistics) to calculate deprivation status.	Year group, type of school/ provision, needs met in provision, SEN status, free school meal (FSM) eligibility.	Primary reason for placement (mental health needs, permanent exclusion, young offender institute, etc.)	EHCP request (date, outcome, tribunal and mediation requests).	Type of disability (including learning disability).
Clinical Subset 296 items including intellectual disabilities and co-morbidities (conditions, technology dependencies, family-reported issues).	Pupil SEN type primary and secondary (MLD: Moderate learning difficulty, SLD: Severe learning difficulty, PMLD: Profound and multiple learning difficulties).	Type of setting (non-maintained further education provision, one-on-one tuition, work-based placement etc.)	SEN settings, placements, type of SEN need, annual review.	Factor identified at the end of a social care assessment relating to a child's learning disability and 40 other factors.
Environmental issues, family, schooling, housing, equipment, leisure.	Fixed period (suspension) and permanent school exclusion and reason(s).	Association (whether the placement is associated with any other services that are not education).	EHCP needs assessments EHCP budgets, outcomes, work-placed learning activities.	

2.2. Project two

This service evaluation explores the needs, pathways, risk and protective factors of South Tyneside and Sunderland children suspended or permanently excluded from school between 2019 and 2024.

2.2.1. Project two: aims and objectives

Aims

1. Conduct a service evaluation to identify when and where assessment and identification of SEND takes place with a population of children and young people with SEND excluded from school
2. Investigate if health, education and children's social care services are assessing and identifying the SEND needs of children to reduce school expulsions and suspensions

Objectives

1. Develop an informed understanding of the emerging needs, risk and protective factors for school exclusion in the age ranges 0-5 and 11-15 years of age
2. Determine whether assessments or referrals are being requested by families and the outcome of these
3. Determine how far services (health, education, children's social care) consistently identify the underlying needs of children with SEND and at what stage
4. Analyse if the right services are being provided at the right time for children and young people who have been excluded from school
5. Determine who is referring/supporting/assessing children with SEMH (ADHD) under six years of age
6. Propose improvements to commissioned pathways and services to reduce school exclusion
7. Provide recommendations for early identification of SEND to reduce school exclusions over the longer term

2.2.2. Project two: the datasets

The following administrative datasets are included for project two (Table 2).

Table 2. Health, education and children's social care datasets: project two

Education data sets
School Census: FSM, primary SEN, SEND status and provision, suspensions and permanent exclusions, attendance, attainment, phonics, reading, mathematics, writing.
SEN2: EHCPs, requests and outcome of assessments, mediation, tribunal, EHCP start date, ceasing of EHCP, residential status, SEN setting, work-based learning
Children's social care data sets
CiN census: Ethnicity, type of disability, children in need codes, child protection plans, Section 47 investigations, type of abuse
Children Looked After (SSDA903 Collection): Ethnicity, care status, SDQ scores, accommodation, placement
Health data sets
Primary Care: GP data
Mental Health Services Data Set (MHSDS): records of assessment and services to support children with a mental health condition and/or a need for support with their mental wellbeing and/or a learning disability and/or Autism or any other neurodevelopmental condition
Improving Access to Psychological Therapies (IAPT) services: NHS Talking Therapies Patient level, output based, secondary uses data set which aims to deliver robust, comprehensive, nationally consistent and comparable information for patients accessing NHS-funded IAPT services in England
Mortality Data Set: Death registrations and statistics on causes of death, mortality rates etc.
Accident and Emergency (A&E): This comes from the Secondary Care Uses (SUS) data warehouse containing details of A&E attendances at National Health Service (NHS) hospitals in England. The Emergency Care Data Set (ECDS) is the national data set for urgent and emergency care. It replaced the HES, A&E dataset used to collect information from emergency departments
Children and Young People Services (CYPS)
Inpatient: The data source for this are Hospital Episode Statistics (HES) and each episode relates to a period of care for a patient under a single consultant at a single hospital. It shows the number of episodes during the period, with breakdowns including by patient's age, gender, diagnosis, procedure involved and by provider.
Outpatient: The data source for this is also HES, containing details of all admissions and outpatient appointments at National Health Service (NHS) hospitals in England. It includes private patients treated in NHS hospitals, patients who were resident outside of England and care delivered by treatment centres (including those in the independent sector) funded by the NHS.
Births/deaths: From local registries.

From October 2017, the Children and Young People's Health Services Data Set was retired and incorporated into the all-age Community Services Data Set (CSDS). The CSDS is a patient-level, output-based, secondary uses data set which aspires to deliver 'robust, comprehensive, nationally consistent and comparable person-centred information for people who are in contact with publicly funded Community Health Services' (NHS, 2023a).

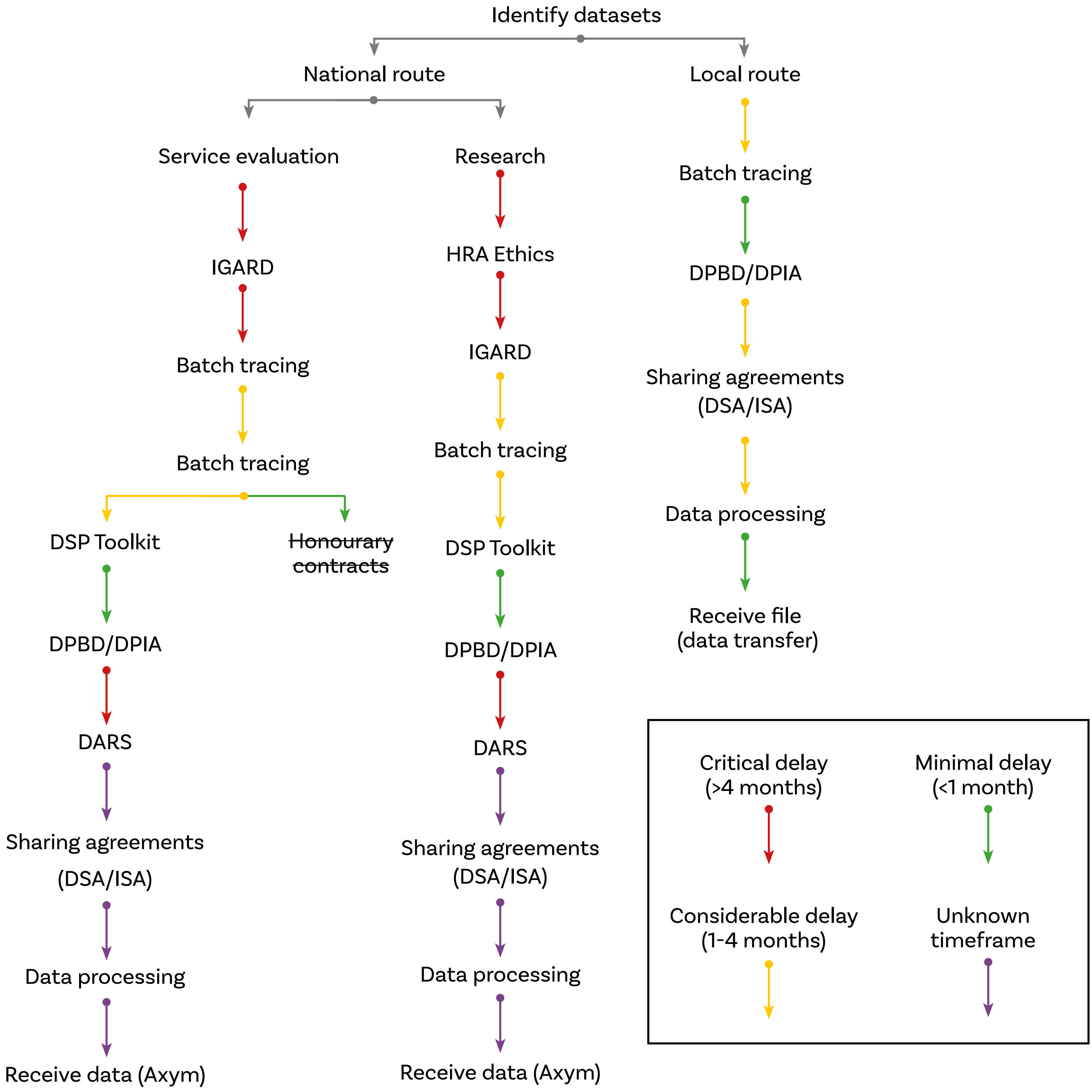
3.0 The data linkage processes for the projects

Section 3.1 provides an overview of the potential data linkage routes for projects one (local route) and two (national route). 3.2 details the legal basis and GDPR rationale and approach. Sections 3.3 to 3.9 share the different processes followed to gain the necessary permissions for each project.

3.1. Data linkage routes

Two possible data linkage routes were available for the two projects; local and national. The local route would require data sharing agreements (DSAs) with TfC, STSFT and NECS. The national route required additional processes as shown in Figure 1. Both projects were undertaken as service evaluations, since a 'research' project through the national route requires an HRA ethics application. Information on whether a project falls under research or service evaluation can be found here (Bristol University Hospital, 2022).

Figure 1. Flowchart illustrating the various IG routes and processes navigated during the two projects



Initially, as the projects were running concurrently and following meetings with local partners in NECS, TfC, Sunderland City Council (SCC), South Tyneside Council, STSFT and the UoS it was decided the national route would be most appropriate for both projects. This was due to the number of health datasets to be included for project two that could only be accessed via NHS England's IG processes.

As project one only required the Community Services Data Set CSDS from the NHS it was decided that this could be achieved through a DSA directly with STSFT and the UoS. This arrangement was viable due to strong local partnerships between the project partners. The following sections outline the IG process for both projects, and indicates whether each process was required for the national/local route, for example:

National route: Required

Local route: Not required

3.2. Setting out the legal basis and GDPR

National route: Required

Local route: Required

As of 2018, all research in the UK (as well as other member states) must adhere to the General Data Protection Regulation (GDPR). These regulations were implemented to standardise data privacy laws in its member states, which consist of 28 nations throughout Europe. GDPR was enacted before the UK left the European Union in 2021. Therefore, it is considered a valid UK law.

GDPR dictates how organisations can use personal data. For the UoS to undertake a service evaluation or research project using individual-level data, there must be an appropriate legal basis under GDPR as to why sharing data is fair and lawful. The following sections sets out the legal basis for projects one and two.

3.2.1. The Children and Families Act (2014), Section 22 (a) and (b)

The Children and Families Act is concerned with the identification of children and young people with special educational needs and disabilities

A LA in England must exercise its functions with a view to securing that it identifies:

(a) all the children and young people in its area who have or may have special educational needs, and

(b) all the children and young people in its area who have a disability.

3.2.2. GDPR and applied public interest (Article 6) and significant public interest (Article 9)

The significant public interest (Article 9) was referenced given the nature and scope of the research project. References were also made to condition 16(1a-e, 2a-c, 5a-b) of the Data Protection Act (2018) relating to '*Support for individuals with a particular disability or medical condition.* '

Article 6: Lawfulness of processing

1. Processing shall be lawful only if and to the extent that at least one of the following applies:
    - (c) processing is necessary for compliance with a legal obligation to which the controller is subject.
    - (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.

Article 9: Processing of special categories of personal data

1. Processing of personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation shall be prohibited.
2. Paragraph 1 shall not apply if one of the following applies:
    - (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 Union or Member State law or pursuant to contract with a health professional and subject to the conditions and safeguards referred to in paragraph 3.

While this process allowed the project members to establish legal basis for processing and analysing the data, there were still steps that had to be taken to ensure the processes remained within GDPR compliance. The most crucial of these steps was the pseudonymisation of any identifiable data. Pseudonymisation involves taking an identifiable piece of information about a data subject and replacing this with an anonymised key. This key can be used to link separate data fields for the same subject, but cannot be reversed to obtain the original piece of identifiable information. In the present projects, pseudonymisation had to be applied to NHS numbers, in order to prevent identification of data subjects and their health data.

3.3. Data Protection by Design (DPBD) and Data Protection and Impact Assessment (DPIA)

National route: Required

Local route: Required

Both the DPBD toolkit and the Data Protection and Impact Assessment (DPIA) were required to be submitted to the UoS data protection officers. The DPBD toolkit screens projects for any risks to personally identifiable information in any data being used. The concept of DPBD is required by GDPR to be built into any accordant organisation, and the UoS uses this toolkit to ensure that any studies comply with GDPR regulations through the whole life cycle of the project.

One of the purposes of the University's DPBD toolkit is to establish whether the project will involve the use of personal identifiable information (PII). If so, the data protection officer will request that the researchers submit a DPIA form. Completing a DPIA is a process that involves identification and minimisation of data protection risks. Much like a regular risk assessment, a DPIA reports the likelihood of each potential risk materialising, and the impact that this would have on the project staff and/or the data subjects. While the use of PII is one reason that a DPIA may be required, this may also be the case where there are any other 'high risk' factors affecting the data. More guidance on the DPIA and types of data processing that may require a DPIA can be found here on the Information Commissioner's Office website (ICO, 2025).

Whilst the due diligence processes detailed above provided a solid legal basis within the requirements of the General Data Protection Regulations, consideration was also needed in relation to the common law duty of confidentiality. The common law duty of confidentiality requires a clear legal basis (not to be confused with the legal bases within GDPR) for sensitive information to be shared beyond the service provider initially collecting the sensitive personal information. Given the legal bases that are available to set aside the common law duty of confidentiality are very limited, consideration should be given to the pseudonymisation of personally identifiable information to be undertaken at the service provider. By providing pseudonymised information to the service evaluation partner without the pseudonymisation key being shared allows the data to considered anonymised by the service evaluation partner therefore descoping the shared data from the requirements of the common law duty of confidentiality.

3.4. Independent Group Advising on the Release of Data (IGARD) Application

National route: Required

Local route: Not required

At the start of the projects, none of the partners were aware that we would need IGARD approval. IGARD were an advisory body in NHS Digital Board, tasked with improving transparent, accountability, quality and consistency of access to NHS Digital Data. NHS digital guidelines stated that NECS, as data processors, couldn't link health, education and care data without IGARD approval. As a collaborative team, the IGARD application was written and submitted in July 2022, to show that the NHS Digital safeguards were met (using a controlled environment to minimise risk of disclosure, to evidence a health benefit and how sharing the data was in the public interest).

Initially, it was believed that IGARD would decide whether to grant permission for NECS to link the LA and health datasets. It was later determined that IGARD only reviewed the application and provided recommendations to NHS Digital on whether to provide access to the healthcare data. This recommendation was received by the project team March 2023, however, IGARD ceased to exist on 31 January 2023 when NHS Digital merged with NHS England. It has since been replaced by the Advisory Group for Data (AGD) which now provides 'expert advice' to the NHS on internal and external data access (NHS, 2025).

3.5. Data Security and Protection Toolkit (DSPT)

National route: Required

Local route: Not required

For organisations to be given data controller or processor eligibility with regards to patient data from the NHS, they must complete DSPT accreditation annually. The DSPT is an online self-assessment tool that allows organisations to measure their performance against the National Data Guardian's 10 data security standards. The self-assessment requires considerable time to complete and must be completed by the researchers as well as the organisation's data protection administrators. It should also be noted that completion of DSPT accreditation is mandatory before submitting an application to DARS.

Due to delays with IGARD and DARS, the use of honorary contracts between UoS and SCC was explored, as the latter was DSPT-certified. Consideration was given to SCC acting as guarantor through a 'parent company guarantee'. Eventually, all parties decided this was not in the interests of either organisation due to risk. Exploring this approach took over three years before being discounted. This meant that the UoS would have to complete DSPT certification themselves, which would delay the projects by at least two months. At this point, to avoid this additional delay, the two projects diverged as UoS elected to pursue a local data linkage route for project one, rather than the national route, which would eliminate the need for DSPT accreditation. DSPT accreditation was eventually obtained by the UoS for project two and has been successfully renewed each year since 2023.

3.6. Batch tracing

National route: Required (if NHS numbers not attached to LA data)

Local route: Required (if NHS numbers not attached to LA data)

The first major concern about the quality of the linked data was that not all children and young people's data could be linked. If individuals' data could not be linked across datasets using a unique child identifier, this would exclude them from analysis, which would have implications for the reliability of the data. The project team decided early in the projects to use the NHS number as the identifier.

Children and young people's NHS numbers are often not known to the LA unless they have taken steps to record them on local data systems (Council for Disabled Children, 2021). In South Tyneside and Sunderland, only CiN census data may contain numbers held locally, not education datasets. The NHS Digital's Demographic Batch Service can perform a one-off batch-tracing exercise. In this context, batch tracing refers to locating NHS numbers based on other demographic information about the individual in LA datasets. However, this does not provide 100% matches due to incompatible addresses or demographic information, such as recorded gender or ethnicity. Through batch tracing, it was evident that caregivers may update their address with their GP, but not with their school. Additionally, 'double-barrelled' surnames may be inconsistently recorded. Discrepancies like these may prevent a child's data from being linked between different services via batch tracing. The success rate in batch tracing for projects one and two was over 90% (highly satisfactory) for different data sets. It should also be noted that the process of batch tracing requires a contract with the data processor (in this instance, NECS).

3.7. Data Access Request Service (DARS)

National route: Required

Local route: Not required

DARS is the gateway to accessing NHS data through various methods such as:

- Secure Data Environment (SDE)
- Clinical trials
- Secure electronic file transfer

The DARS application for project two was submitted October 2023, the outcome of this remains unknown twenty months on. There are no timescales available for the DARS completion or outcome and this continues to be the barrier to project two progressing. No data can be shared by NECS until this is approved by the Department of Health and Social Care.

The data sets available through DARS are limited to those found in their data set catalogue. Prior to applying for any of these data sets, the applicant should check the pre-application checklist, which includes other processes discussed in this briefing paper such as the DSPT and valid legal basis.

3.8. Data sharing agreements (DSAs)

National route: Required

Local route: Required

DSAs are a common pre-requisite for any personally identifiable information to be shared between agencies / organisations. A DSA allows for numerous organisations to clarify who is responsible for meeting accountability obligations under the UK GDPR (ICO, 2024a), or any other data protection and privacy legislation currently in effect under law.

One of the fundamental requirements for the development of a sharing agreement is to ensure there is a detailed understanding of:

- Which individuals are subject to data sharing
- What data is to be shared, as detailed a specification as possible
- How the data will be shared
- Which organisations is the data to be shared with, and in what format
- What organizational and technical controls are in place to ensure the ongoing protection of data

A DSA can exist between a data controller and data processor, or between two data controllers. In the case of a controller to processor agreement, the DSA states the agreed purpose of the data processing and what processing will be carried out by the data processor. A dual controller DSA entails each controller assisting the other in complying with all applicable requirements of the current data protection legislation.

Information Commissioner's Office – What are data controllers and processors? (ICO, 2024b)

Controllers make decisions about processing activities. They exercise overall control of the personal data being processed and are ultimately in charge of and responsible for the processing.

Processors act on behalf of the relevant controller and under their authority. In doing so, they serve the controller's interests rather than their own.

To minimise delays in the preparation of sharing agreements, it is important that the controllers of the datasets that are subject to sharing are the key developers of the agreements, as the controlling organisations set the requirements of the processors. The processing party should ensure that they are amenable to all requirements within the agreement prior to signing. To assist with this, relevantly skilled IG contacts from each controlling/processing organisation should meet to develop the agreements and continue to meet until agreements are signed.

As part of project one and two there were multiple modifications to approved DSAs, due to:

- Removing and adding data sets
- Changes in age range of children and young people to reduce workload on LAs (see section 3.11)
- Changes in designation of organisations (processor or controller)
- Changes in how data was to be shared between organisations
- Changes in who would share the data between organisations

3.9. Data processing

Data processing is the act of taking any action on personally identifiable information, be that collection, storage, or manipulation up to and including the point of destruction. It should be noted that data controllers determine how data will be processed and may outsource or pass some of these processing activities to a data processor. In the event that a data processor is used the data controller is required to ensure that a robust agreement and contract exists between themselves and the processor even where the processor is in receipt of data for the purposes of data evaluation or research activity, and in the case of NHS data even where the data being provided to the processor is in anonymised form.

The UoS sought pseudonymised data for the purposes of both service evaluations. In their role as data processors, NECS were commissioned to support the IG for both projects and to pseudonymise the health data prior to sharing with the UoS. This involved NECS replacing the NHS number across health datasets using an unreversible pseudonymisation key (Pseudo@ Source tool). Where a sensitive or confidential code is found, all coding for that patient on that date is removed from the pseudonymised dataset.

NECS also supported TfC and South Tyneside Council in preparing the LA data in the use of the Pseudo at Source tool. Through DSAs the LA then transfer datasets that have been through the Pseudo@Source tool to ensure a match with the already pseudonymised health data received by the UoS. This was to ensure the UoS would be able to link the data within their secure environment with the pseudo code for individual children.

3.10. Receipt of data

Whilst agreement may be reached on the types of data to be shared it is an important step to ensure that all parties know how to send and receive data to and from each other through a DSA. For project one, several DSAs were agreed to ensure the safe sending and receiving of data between:

- TfC and UoS
- TfC and NECS
- NECS and UoS

It is critical that whilst the data is transferred between one party and another it continues to require a level of protection in keeping with its sensitivity. It is often helpful for the recipient organisation to develop a process whereby the controlling organisation is provided with temporary access to processors systems directly to deposit the data in question thus removing the requirement for the data to be sent either by email or by using third party data transfer services. Where this is not possible it is beneficial for the data controller to identify a third-party transfer service that they are comfortable using to ensure additional checks are not required to be undertaken around the security of the data transfer service, where that service is proposed by the processor/recipient of the data.

3.11. Insights from discussions with project partners

Capacity in LAs is a limiting factor in pursuing data linkage projects. The time taken to navigate the IG, prepare, pseudonymise and share data requires commitment and a drive to be innovative. TfC identified the potential benefits on outcomes for children and families in learning from linked data and this drove the projects forward.

One of the barriers encountered in the project was that, although IG processes had been completed, the LAs had to process their data in such a way that it complied with GDPR and the DSAs for transfer to the researchers. For example, the datasets could not be transferred to the researchers with identifiable information included for project one (this would have required the completion and approval of a DPIA). This meant that, either prior to pseudonymisation or transferring the data to the UoS, the LA would have to remove identifiable information, such as names and addresses. The extent to which this could impact a data linkage project depends on the number of data files being transferred. For project one, the LA was attempting to transfer 79 data files, and removing identifiable information from each individual in these datasets considerably added to the workload. A similar barrier had been previously encountered, in which TfC did not have the capacity to remove children of certain ages from the datasets, however this difficulty was overcome by modifying the DSA to allow the LA to include children and young people of all ages (0-25) in the dataset (as opposed to ages 5-18).

One of the explanations for the high capacity required on behalf of data controllers such as TfC is the way in which LA data is collected and stored. TfC explained that the number of datasets was higher than expected because different years of data could not be merged into one dataset. This was attributed to the changeable nature of terminology in children's social care and education data collection.

3.12. Barriers and enablers reported by local authorities: national research

A publication completed as a part of these data linkage projects was a policy brief detailing insights from interviews with LA staff across England. Some of these insights overlap with experiences from the data linkage projects, however there is valuable knowledge exchange to be had from these interviews, as they contain novel perspectives and experiences (Pinney et al., 2024).

Local authorities reported several barriers in linking administrative data on children and young people with SEND. A key challenge is the complexity of health data, including limited knowledge among LA partners and a perceived reluctance from health services to share information due to risk aversion. Information governance requirements add further complications, with time-consuming processes for securing data sharing agreements. Additionally, technical hurdles such as incompatible IT systems, inconsistent identifiers, and poor data quality hinder efficient data linkage, as has been the experience across the projects. Cultural challenges also play a role, as services traditionally work in silos and may resist collaborative data sharing efforts, particularly when dealing with backlogs and resource constraints.

Despite these challenges, the local authorities reported several enabling factors that support progress in data linkage. Strong senior leadership and governance structures were found to be crucial, and SEND data is often given strategic attention and drives multi-agency collaboration. Where present, integrated teams and shared systems help overcome technical and organisational barriers by providing a streamlined approach to data collection and usage.

Local areas that had prior experience in data linkage benefited from existing expertise and infrastructure to navigate challenges effectively. Additionally, having professionals with NHS knowledge within LA teams proved invaluable in understanding health data and fostering partnerships across sectors (Pinney et al., 2024).

4.0 Conclusion

There continues to be no national guidance on best practice processes for linking data across agencies. This would have been helpful in preparing the partners for compliancy expectations. Knowing that access to health datasets can only be achieved via mandated NHS IG processes, including the HRA, IGARD, DSPT, and DARS would allow for preparation and feasibility costing as part of project design.

The principal difficulty in completing the IG for this project was that systems and processes related to data protection were not sufficiently clear or consistent. Despite a great deal of time and effort, as well as the collective expertise of the project partners, new requirements suddenly emerge, or requirements and organisations change, creating new hurdles or removing obstacles after the team had spent months attempting to navigate them. This impact should not be underestimated: sending and receiving data cannot proceed until all the IG is agreed upon and approved. In the meantime, funders' deadlines may expire, team members may change roles, other priorities may come to the fore and project partners' ability to continue may be eroded.

The IG procedures accounted for most of the difficulties and barriers experienced during the data linkage projects. Significant time and effort were needed from all parties, this was difficult even with strong partnerships in place. Partnership working requires a high degree of commitment, which may be eroded with each additional delay. Delays were common regardless of following the local or national route. Often this was due to pressures on services, staff capacity, funding, staff changes, annual leave or a lack of control over process timeframes (e.g. IGARD and DARS) or knowing the processes to follow.

The lack of a consistent child identifier makes it challenging to link data sets efficiently, securely and with a high degree of confidence or accuracy. The fact that the NHS number is not included in education datasets was the most significant challenge to overcome. However, there is optimism about piloting a single unique identifier in the future. Involving fewer data sets, fewer partners or limiting to a small number of data items at key points in time (i.e. 'snapshots' of key indicators) eases the IG. However, the hope is that we can learn from these projects to improve processes that utilise linked data, thereby enhancing the lives of children and families.

Future projects may have greater success navigating barriers if healthcare, social care and education terminology are standardised between census years, and across services. However, while the variability of data fields and terminology can act as a barrier for data linkage, forced standardisation and constraints may discourage innovation, and hinder local authorities in their ability to capture greater and richer detail in their data collection approaches.

It has been five years since the project team and partners began the projects. On publication of this briefing paper, the pseudomymised datasets across health, education and children's social care have been received for project 1, and now analysis and learning can begin. Project two has been held up by DARS indefinitely, but a recent development has allowed for a sub-licensing agreement with the North East and North Cumbria ICB, allowing the sharing of health data between them and TfC and mitigating the need for DARS approval. We hope to resume collaboration with local partners and link these additional heal

5.0 Recommendations

Recommendation 1	Use a consistent, unique identifier, such as the NHS number, to support efficient and effective data sharing and linkage.
Recommendation 2	Establish a national mandate and funding to link the health, education and social care data of children with SEND, with clear expectations on data sharing between health and LA services through compatible data systems.
Recommendation 3	Develop a joined-up approach from central Government and the inspectorates, so local areas can be confident that their investment in developing data linkage systems will meet expectations.
Recommendation 4	Fund opportunities for LAs and health partners to share learning about how they have successfully linked health and LA data to improve outcomes for children.
Recommendation 5	Invest in developing best practice guidance and IG templates to make it easier for ICBs and LAs to develop data sharing protocols and processes.
Recommendation 6	Invest in a consistent and easy to use data capture interface to improve documentation in electronic medical records of the multifaceted needs of everyone attending a clinical encounter in the health service, regardless of setting. This should be underpinned by a harmonised set of agreed terms linked to SNOMED-CT codes so that reliable comparisons about all needs can be made over time and across settings. These data would be stored in secure data repositories in NHS organisations and be available to inform service design to meet all needs efficiently and effectively, quality improvements, research and for secondary uses such as reporting to NHS mandated data sets, which currently include the Community Services Data Set, a key dataset for understanding trends in SEND and the multi-faceted needs of disabled children and their families.
Recommendation 7	Mandate consistent data capture and documentation of all needs at all points of care and services across agencies to ensure that quality data are available to link and analyse to improve outcomes for children and young people, with a transparent process of consent in place for documentation, storage and use of these data.
Recommendation 8	Clarify the definitions of the different terms used across services and agencies to describe learning disabilities and learning difficulties and the assessment pathways that underpin these.

References

ADR UK (2023) *Linked local data on children and young people*. Available at: https://www.adruk.org/ our-work/browse-all-projects/ linked-local-data-onchildren-and-young-people/.

Aldridge, R.W., Menezes, D., Lewer, D, Cornes, M., Evans, H., Blackburn, R.M., Byng, R., Clark, M., Denaxas, S,. Fuller, J., Hewett, N., Kilmister, A., Luchenski, S., Manthorpe, J., McKee, M., Neale. J., Story, A., Tinelli, M., Whiteford, M., and Wurie, F. (2019) 'Causes of death among homeless people: a population-based cross-sectional study of linked hospitalisation and mortality data in England', *Wellcome Open Research*, 4(49), 1-17.

Atherton, I.M. *et al.* (2015) 'Barriers and Solutions to Linking and Using Health and Social Care Data in Scotland', *The British Journal of Social Work*, 45(5), pp. 1614–1622. Available at: https://doi.org/10.1093/bjsw/bcv047.

Children and Families Act 2014 (2014) Section 22. Available at: https://www.legislation.gov.uk/ukpga/2014/6/section/22

Chiu, M., Lebenbaum, M., Lam, K., Chong, N., Azimaee, M., Iron, K., Manuel, D. and Guttmann, A. (2016) 'Describing the linkages of the immigration, refugees and citizenship Canada permanent resident data and vital statistics death registry to Ontario's administrative health database', *BMC Medicine Informing Decision Making,* 16(135), 1-11.

Council for Disabled Children (CDC) (2021) *Putting the NHS Number on an EHC Plan and making a request for access to the NHS Digital Personal Demographics Service*. Available at: https://councilfordisabledchildren.org.uk/resources/all-resources/filter/health/putting-nhs-number-ehc-plan.

Department for Education (DfE) (2022) *SEND Review: Right support, Right place, Right time*. London. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1063620/SEND_review_right_support_right_place_right_time_accessible.pdf.

DfE (2023) *Improving multiagency information sharing - Government policy on information sharing and the use of a consistent child identifier*. Available at: https://www.gov.uk/government/publications/improving-multi-agency-information-sharing.

DfE (2025) *Children's Wellbeing and Schools Bill*. Available at: https://bills.parliament.uk/bills/3909 (Accessed: 16 June 2025).

Downs, J., Gilbert, R., Hayes, R.D. et al. (2017) 'Linking health and education data to plan and evaluate services for children', *Archives of Disease in Childhood,* 102, 599-602.

Downs, J.M., Ford, T., Stewart, R., Epstein, S., Shetty, H., Little, R. and Hayes, R. et al. (2019) 'An approach to linking education, social care and electronic health records for children and young people in South London: a linkage study of child and adolescent mental health service data', *British Medical Journal*, 9(1), 1-12.

Fleming, M., McLay, J.S., Clark, D., King, A., Mackay, D.F., Minnis, H and Pell, J.P. (2021) 'Educational and health outcomes of schoolchildren in local authority care in Scotland: A retrospective record linkage study', *PLoS Medicine*, 18(11), 1-20.

Government Actuary's Department (GAD) (2021) *Mortality Insights from GAD - September 2021, GOV.UK*. Available at: https://www.gov.uk/government/publications/mortality-insights-from-gad-september-2021/mortality-insights-from-gad-september-2021 (Accessed: 18 June 2025).

Health and Care Act 2022 Part 6, S.175 - *https://bills.parliament.uk/bills/3022*

Harron, K. (2022) 'Data linkage in medical research', *BMJ Medicine*, 2(1), 1-3.

Holman, C.D.A.J., Bass, J.A., Rosman, D.L., Smith, M.B., Semmens, J.B., Glasson, E.J. and Stanley, F.J. et al. (2008) A decade of data linkage in Western Australia: strategic design, applications and benefits of the WA data linkage system. *Australian Health Review*, 32(4), 766-777.

Horridge, K. (2018) *SNOMED CT Case study*. London: Council for Disabled Children and NHS. Available at: https://councilfordisabledchildren.org.uk/sites/default/files/uploads/files/SNOMED-CT%20Case%20Study%20-%20Sunderland.pdf

Information Commissioner's Office (2024a) *Data sharing agreements*. ICO. Available at: https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/data sharing/data sharing-a-code-of-practice/data sharing-agreements/ (Accessed: 22 April 2025).

Information Commissioner's Office (2024b) *What are 'controllers' and 'processors'?* ICO. Available at: https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/controllers-and-processors/controllers-and-processors/what-are-controllers-and-processors/ (Accessed: 14 April 2025).

Information Commissioner's Office (2025). *Data Protection Impact Assessments (DPIAs).* Available at: ICO website [Accessed 9 June 2025]

Krasniki, P., Carr, M and Stevens, M. (2023) *"People don't understand" - The impact of the cost-of-living crisis on children and young people with special educational needs and disabilities.* London: The Childhood Trust.

Labour Party (2024) *Mission 5: Breaking down barriers to opportunity* (https://labour.org.uk/wp-content/uploads/2023/07/Mission-breaking-down-barriers.pdf)

Long, R., Roberts, N. and Foster, D. (2025) *Children's Wellbeing and Schools Bill 2024-25.* House of Commons Library.

Martin-Denham, S., Scott, N., Horridge, K. and Pinney, A. (2023) Putting together the data jigsaw: The extent of the linking of administrative datasets on children by local authorities in England. Responses to a Freedom of Information request. Discussion Paper. University of Sunderland, Sunderland.

NHS (2025) *Advisory Group for Data (AGD), NHS England Digital.* Available at: https://digital.nhs.uk/about-nhs-digital/corporate-information-and-documents/advisory-group-for-data (Accessed: 18 June 2025).

Parish, N., Bryant, B. and Swords, B. (2019) *Have we reached a 'tipping point'? Trends in spending for children and young people with SEND in England.* Local Government Association. Available at: https://www.local.gov.uk/have-we-reached-tipping-point-trends-spending-children-and-young-people-send-england.

Parson, S. and Platt, L. (2022) *Special educational needs and disability: a lifetime of disadvantage in the labour market? CLS working paper 2022/24.* London: University College London.

Pinney, A., Martin-Denham, S., Horridge, K and Scott, N. (2024) *Putting together the data jigsaw: Linking administrative data sets on children with SEND - Learning from local area experience.* University of Sunderland, Sunderland.

Scott, N., Martin-Denham, S., Horridge, K. and Pinney, A. (2024) *Putting together the data jigsaw: The extent of the linking of administrative datasets on children by integrated care boards in England. Responses to a Freedom of Information request.* University of Sunderland, Sunderland.

Sohal, K., Mason, D., Birkinshaw, J., West, J., McEachan, R.R., Elshehaly, M. and Wright, J. et al. (2022) Connected Bradford: A Whole System Data Linkage Accelerator. *Wellcome Open Research*, 7(26), 26.

University Hospital Bristol (2022) *Is it research, audit or service evaluation?* Available at: https://www.uhbristol.nhs.uk/research-innovation/for-researchers/is-it-research,-audit-or-service-evaluation/ (Accessed: 9 June 2025).

Wilcock, R., Elliot, H. and Symons, T. (2022) *Unlocking the potential of early years data.* Available at: https://www.nesta.org.uk/blog/unlocking-the-potential-of-early-years-data/ (Accessed: 9 May 2025).