# Data Integration Platform – Learning Journey

LEAP

# Table of Contents

# 1.0 Purpose of this report

The concept of a Learning Journey was developed to capture the story of a LEAP service over its lifetime: from its inception to the end of its journey as part of the LEAP programme.

This report is different from LEAP's other Learning Journeys in that it is not service-focused. Instead, it focuses on the development and implementation of LEAP's Data Integration Platform ('The Platform'), which was a crucial component of the programme's routine Monitoring, Evaluation and Learning approach throughout its second half (2020-2025).

This report and LEAP's other Learning Journeys will be shared with key stakeholders including the National Lottery Community Fund, local early years commissioners and public health colleagues, service delivery partners, families and via national networks. We hope that learning from the LEAP programme can inform future commissioning and programming decisions and contribute to the wider evidence base about health improvement interventions in the earliest years.

This report is complemented by a learning journey focused on LEAP's Shared Measurement System (SMS) (https://leaplambeth.org.uk/reports/shared-measurement-system-learning-journey/). The SMS was developed at the same time as the Data Integration Platform and acted as the foundation for LEAP's new approach to routine Monitoring, Evaluation and Learning.

Where the SMS provided a new, theory-informed framework for data collection, The Platform was the mechanism that allowed this data to be processed and analysed, linking individual-level data across services and administrative datasets to provide a programme-wide view of LEAP's work.

The SMS Learning Journey is referenced where appropriate throughout this report.

## Authors

Rowan Ferguson, Senior Research and Evaluation Officer, National Children's Bureau
Sophie Woodhead, Assistant Director, Lambeth Early Action Partnership
Gemma Luck, Public Health Intelligence Manager, Lambeth Early Action Partnership
David Wood, Programme Manager, Lambeth Early Action Partnership

## About Lambeth Early Action Partnership

Lambeth Early Action Partnership (LEAP) was one of five local partnerships which made up A Better Start (ABS). ABS was a national ten-year (2015-2025) test and learn programme funded by the National Lottery Community Fund that aimed to improve the life chances of babies, very young children, and families. LEAP was supported by the National Children's Bureau charity (NCB) and worked with a wide range of families, practitioners, and organisations across Lambeth.

LEAP's aims were to:

- Improve early child development outcomes for all children living in the LEAP area
- Reduce local inequalities by supporting those at greater risk of poor outcomes

LEAP operated in a focused area of Lambeth which was selected based on local need, drawing on a range of local evidence that illustrated greater inequalities for young children in this area compared with the rest of Lambeth.

LEAP funded and developed more than 20 local services to meet the needs of families through pregnancy and the early years of childhood. LEAP was a 'collective impact initiative', which means that all of its services and activities linked together and worked towards shared goals to improve outcomes for very young children.

A glossary of key terms relating to the LEAP Programme can be found in Appendix 1.

# 2.0 Background to data integration in the early years

## 2.1 Why is data integration in the early years important?

There is strong consensus that the 1,001 days from pregnancy to the age of two lay the foundations for an individual's physical, cognitive and emotional development over the rest of their life course.[1] Early intervention in this period is also recognised as a key protective factor for children's development[2] and for addressing inequalities[3]. This means helping children and families to access support before an issue escalates.

In this context, the quality of the services that support children and families is essential. Evidence shows that whether services can successfully work together is a crucial part of this.[4]

Joint use of data is one of the key ways that that the early years system can come together to offer high quality support to children and families.[5] This sits alongside other forms of partnership, including joint governance and delivery. There are two main ways that organisations can jointly use data:

1. Data sharing

Organisations provide partners with access to information that they cannot access in their own data systems.

2. Data integration

A more intensive process that involves combining data from multiple sources into a single, unified view using common data fields.[6]

This report is primarily focused on data integration.

The benefits of integrating data are widely documented. It helps organisations become better commissioners, iteratively improve their services, and more effectively meet the needs of children and families. Commonly recognised benefits include:

- **Enabling more informed commissioning**
  Allowing a more complete picture of needs, risk factors and outcomes. Bringing to life the well-recognised links between health, education and social outcomes.[7]

- **Facilitating joined-up care**
  Identifying problems early through improved access to information collected by other agencies (e.g. health conditions, social factors and other key determinants of health). Enabling timely access to crucial safeguarding information. Tracing how services intersect and interact with families, enabling the identification of new opportunities for collaboration and improvement.

- **Reducing the burden on families**
  Ensuring that they don't need to repeat the same information to practitioners on multiple occasions.

- **Understanding what works**
  Identifying whether engagement with certain combinations of services improves outcomes.

## 2.2 The current state of data integration in the early years

Despite these widely recognised benefits, many organisations have noted that data in the early years is typically fragmented and siloed.[8] Though there is no shortage of data being collected, much of it is not shared and utilised across organisations.

There is clear evidence of this in the context of local early years systems. Research by the University of Sunderland revealed that almost a third of local authorities were not linking internally held education and social care data. Of the local authorities who did link these datasets, most also opted to focus on specific groups (i.e. children with SEND).[9]

Integrating health data with data with education and social care data is even more of a problem. This is despite clear evidence of the link between these areas of development. Over three quarters of local authorities did not link their internal data with health datasets collected by organisations in the wider system.[10]

Despite these blind spots, only 6% of integrated care system strategic plans look to address challenges around data and information sharing for babies, children and young people – according to a 2024 review of integrated care systems' strategies.[11]

Therefore, though successful data integration does happen, this is typically not a seamless flow information across systems. Instead, is often incomplete, reliant on dedicated individuals, or done on a case-by-case basis.[5]

### Key challenges around data integration

The sheer number of actors in the early years sector presents an underlying challenge to data integration.[12]

Between the ages of 0-5, a child and their family are likely to interact with numerous services and activities. These are provided by a wide range of statutory, private, voluntary, community or faith-based organisations. Within a short time period, a single child and their family might receive support from an NHS Health Visitor, a Children's Centre, Children's Social Care, a private nursery school, and a community-based food programme.

Integrating data across so many diverse services is fundamentally difficult. This is made more challenging by several other issues:[5][13]

1. A lack of common identifiers across services

This means it is not possible to easily link individuals across datasets. Some subgroups of services do use this approach, i.e. services that use NHS numbers, Unique Pupil Numbers and Local Authority IDs. However, there is currently no widely used system that routinely relates one identifier to another.

2. Most services see themselves as working towards different goals

This hinders the case for the value of integration.

3. Data capacity and infrastructure is highly variable

There are recurring issues around the quality and variability of IT systems, an overstretched and undermanned data-focused workforce, and different interpretations of key legislation, like UK General Data Protection Regulation (GDPR).

4. Developing information governance processes is fraught with difficulties

This includes identifying the right legal basis for data sharing and going through the often-arduous process of establishing data sharing agreements.

All these issues are made worse by chronic underfunding of early years services, pushing data integration towards the bottom of the priorities list.

### Consequences of poor data integration in the early years

As a result of these challenges, even when services are seeing and working with the same families, they tend to collect, store and analyse data separately. Therefore, services and professionals often have an incomplete understanding of a family and child's needs. This makes it difficult for them to provide integrated support that takes a holistic, family- and child-centred approach.[8]

It adds up to a lack of understanding of how different families interact with and move across multiple services. This prevents a full appreciation of what does and doesn't work to improve child and family outcomes.

## 2.3 The UK policy context

### Calls for improvement in data sharing and integration

There is a long history of attempts to improve data sharing and integration around children and families.[5] Take, for example, the early case made by the 2011 Munro Review of Child Protection.[14]

Momentum has grown in recent years. The Children's Commissioner for England , for example, has highlighted missed opportunities for diverse services to use data to understand the interrelated nature of the work. In particular, she spotlighted the potential to use health data to understand issues with school attendance.[15] Similarly, Health Equity North has advocated for data integration as way to speed up and improve coordination of support across public services.[16]

The Academy of Medical Sciences identifies improving the collection, access and linking of early years data as one of five key priorities to support improvements in health and wellbeing, and reduce inequalities in the first five years of life.[17] In particular, it focuses on the importance of collecting and linking data on the wider determinants of health[18] (the social, economic and environmental factors which influence mental and physical health). It argues that the data will help to better understand whether interventions and policies are making a difference. This idea is echoed by the Care Quality Commission in the context of maternity care.[19]

### Examples of successful practice

There are some examples of successful data integration in the early years. University College London's ECHILD Database,[20] for example, brings education data from the National Pupil Database together with health data from the Hospital Episode Statistics database. It aims to understand how children's education affects their health and vice versa.

In a local context, Liverpool's Family Build System,[21] combines data from the Department for Work and Pensions, health, police, education and early help to help social workers judge a child's level of risk. Outside of the early years, the Multi Agency Safeguarding Tracker (MAST)[22] links data from across NHS, social care, and fire and rescue to help professionals make informed safeguarding decisions.

### The move towards a Single Unique Identifier

One key area of policy development are moves to introduce a Single Unique Identifier (SUI) to make data integration across services more straightforward. A SUI uses a consistent code to link individuals across different datasets.

Organisations such as the National Children's Bureau,[23] have been calling for an SUI for more than a decade, with the early years sector commonly felt to be behind other sectors. Some Local Authorities and programmes already have similar systems in place, typically using a mix of demographic information and identification numbers (including the NHS Number) as a shared identifier.[13]

In 2022, the Health and Care Act[24] called for the Government to set out a clear policy on the use of a consistent identifier for children. Following this, and in line with an earlier manifesto commitment[25] in 2024 the Labour Government published 'Keeping Children Safe, Helping Families Thrive', which outlined plans for whole-system reform of children's social care. This included a commitment to the eventual introduction of an SUI to improve multi-agency information sharing.

Additionally, the Children's Wellbeing and Schools Bill[26] is undergoing parliamentary scrutiny at the time of writing. The Act includes measures to improve inter-agency information sharing when relevant to safeguarding or children's welfare. These measures cover requirements around the use of an SUI.

Despite the momentum to introduce an SUI, there is widespread recognition that it is not a silver bullet for improved data integration. Other contextual challenges need to be resolved for an SUI to be used effectively, including:

- improving the interoperability of IT systems;
- clarifying the legal basis for data sharing and data use; and
- changing cultures around the sharing and use of data

A focus on safeguarding also means there is currently little discussion around the potential for SUI to improve research and commissioning through enabling a more complete picture of children's needs, risk factors and outcomes.

# 3.0 The need and rationale for LEAP's Data Integration Platform

## 3.1 LEAP's wider work to refresh its approach to Monitoring, Evaluation and Learning

In its first half (2015-2020), LEAP's primary focus was on developing and implementing a high-quality programme. LEAP's second half (2020-2025) had a renewed focus on outcomes, learning and impact.

The Shared Measurement System (SMS) (https://leaplambeth.org.uk/reports/shared-measurement-system-learning-journey/) and the Data Integration Platform were developed to underpin this new phase, aiming to significantly strengthen LEAP's approach to routine Monitoring, Evaluation and Learning. These two systems were complementary. The SMS provided a new, theory-informed framework for data collection. The Platform provided the mechanism for this data to be processed and analysed.

LEAP's SMS was the foundation for its new approach to Monitoring, Evaluation and Learning. Prior to its introduction, both LEAP's theoretical basis and its measurement approach were relatively underdeveloped. A programme-level Theory of Change and some service-level ones did exist, however there was little theoretical alignment across the programme.

LEAP's portfolio of 25 services were working towards over 200 long-term outcomes. Data collection was also primarily focused on performance rather than outcomes. Limited standardisation across data types inhibited LEAP's ability to conduct accurate monitoring – and to combine data to understand the programme's collective impact.

In response to these challenges, the goal of the SMS was to create and implement a common set of monitoring and evaluation tools that all services would find practical to track progress towards and assess achievement of intended outcomes. In doing so, LEAP aimed to:

- improve its ability to pool information across the programme;
- quickly build a larger evidence base; and
- achieve greater efficiency and consistency across data collection, analysis and learning.

Developing the SMS involved several steps. First, refreshing LEAP's Theory of Change. Next, identifying a set of robust (and where possible standardised) measurement tools that would bring this theory to life. Finally, strengthening services' ability to collect high quality and consistent data.

LEAP's overarching Theory of Change and intended long-term outcomes can be found in Appendix 2. Further detail on the SMS can be found in its Learning Journey (https://leaplambeth.org.uk/reports/shared-measurement-system-learning-journey/).

As explored in this Learning Journey, the Data Integration Platform aimed to bring to life the data that was collected via the SMS. It looked to address issues arising from siloed data by linking individual-level data across services and administrative datasets.

## 3.2 What data did The Platform need to host?

Throughout the programme's lifetime, LEAP collected data at service level, while also monitoring local administrative datasets that included population indicators relevant to the programme's long-term outcomes. The Platform processed and stored both sources of data.

As set out above, data collection at a **service-level** was determined by the Shared Measured System, with standardised measurement tools introduced across the programme. LEAP services collected six key data types:

1. **Input data**
   The financial, human, and material resources used in planning and delivery

2. **User data**
   The characteristics and demographics of people accessing services, standardised via a 'minimum dataset' shared across the programme (see Appendix 3)

3. **Engagement data**
   How many people engaged with services and how often

4. **Feedback data**
   What families and practitioners thought of services

5. **Medium-term outcomes data**
   The medium-term changes LEAP wanted to achieve

6. **Long-term outcomes data**
   The long-term changes LEAP wanted to achieve

Administrative datasets onboarded onto The Platform included:

- Maternity Data (Badgernet)
- Health Visiting data (Carenotes)
- Early Years Foundation Stage Profile
- National Child Measurement Programme
- Children's Centre data (EISi)

## 3.3 Data integration prior to The Platform

Prior to the establishment of the Data Integration Platform, LEAP's use of data was heavily siloed.

On a quarterly basis, each provider would submit an anonymised and aggregated data report. Each report would then be uploaded to a service-specific database. These databases would often take the form of unwieldy excel spreadsheet.

Other than for certain subsets of services, this data was not integrated in any meaningful sense. Without any way to identify the same individual in separate reports (i.e. through a unique identifier) there was no way to link individual-level data across different services, or even over multiple quarters within the same service.

The primary reason for this was the breadth and scope of LEAP's activities. Data was being collected by a wide variety of providers (including childcare settings, statutory organisations and voluntary and community sector (VCS) organisations), with consent given on the basis that no identifiable data would be shared with other organisations. Figure 1 shows the diversity of providers collecting or sharing data.

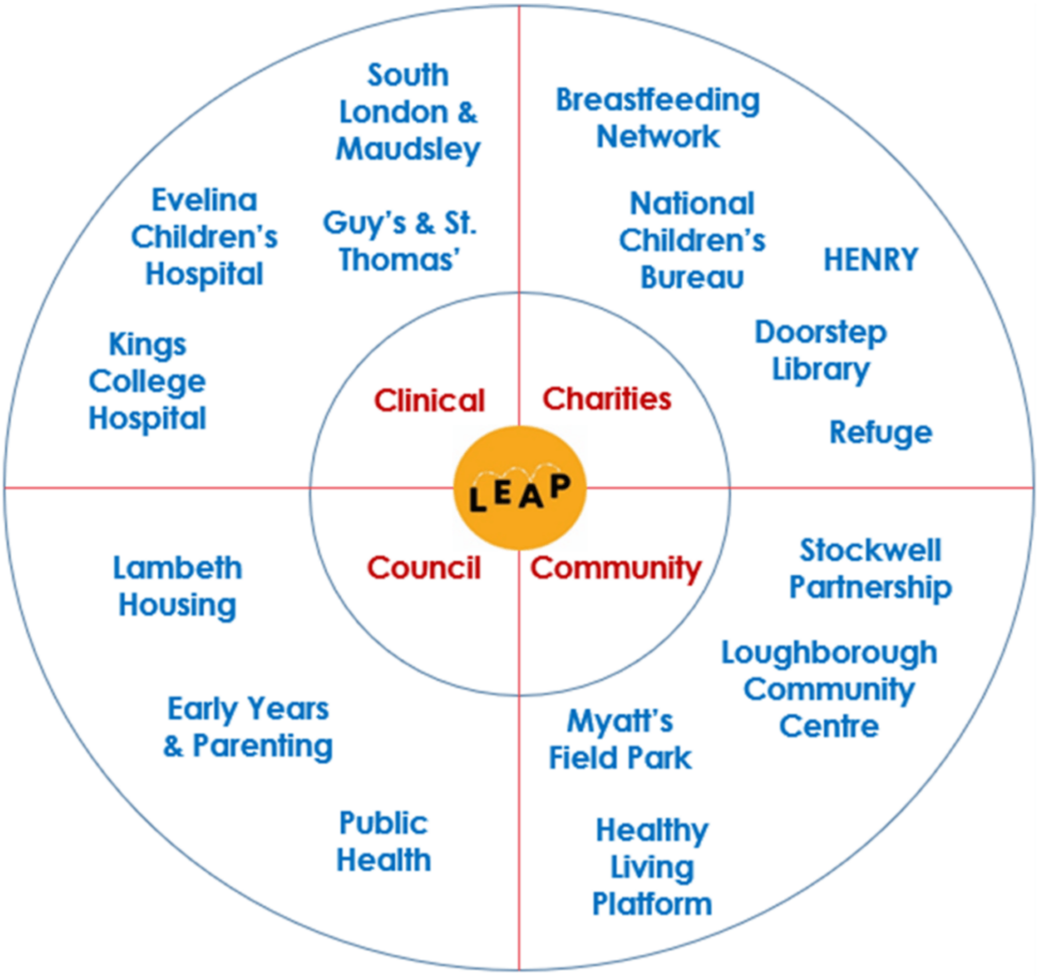


**Figure 1 – The LEAP service providers and statutory partners that collected service-level data, or shared administrative data with the LEAP programme.**

### Consequences of siloed data for LEAP's routine Monitoring, Evaluation and Learning

LEAP's inability to link individual-level data across the programme, or even within one service across reporting quarters, had far-reaching consequences for the programme's routine Monitoring, Evaluation and Learning. Most crucially, it meant all monitoring, analysis and reporting was likely to include double counting.

Double counting prevented LEAP from understanding its work with, or impact on, unique beneficiaries. This meant that the programme was unable to:

- Accurately report on overall reach figures, and the characteristics of those it reached
- Accurately understand patterns of engagement, either with individual services or across multiple services
- Fully understand its collective impact, as it could not isolate multiple-service users to explore the impact on their outcomes
- Link individuals with administrative-level data to understand the programme's long-term impact
- Conduct accurate service-level monitoring due to the potential for double counting

## 3.4 Goals for LEAP’s Data Integration Platform

The Data Integration Platform aimed to enable individuals to be linked across services and administrative datasets, pulling together user, engagement and outcomes data from across all of LEAP’s data sources.

Through doing so, it aimed to:

- bring LEAP’s newly introduced Shared Measurement System to life;
- enable a more holistic understanding of the different factors affecting children’s outcomes; and
- allow for accurate monitoring and reporting.

Specific goals included:

### Capturing programme-wide reach and user characteristics

LEAP wanted to build a comprehensive understanding of who accessed its services (and who did not). It therefore needed to avoid duplicate counting by integrating service-level datasets.

### Understanding LEAP’s collective impact

LEAP aimed be able to understand patterns of engagement with multiple services, as well as the cumulative impact of multiple service use on intended outcomes. To do this, the programme needed to link individuals across datasets.

### Accurate monitoring and reporting across LEAP’s key data types:

LEAP wanted to ensure that service-level reporting included only unique beneficiaries. This would allow for accurate reporting across LEAP’s newly standardised service-level data types. The outputs of The Platform would also reflect data back to service teams in a systematic and clear way – fulfilling a commissioner-level objective.

### Linking individuals to administrative datasets

The Platform also intended to house administrative datasets. These could then be linked to LEAP individuals whose user data was in The Platform, providing a key source of long-term outcomes data. This would provide a far more accurate understanding of LEAP’s long-term impact compared to reporting on data for the LEAP area as a whole.

### The Platform – not a case management system

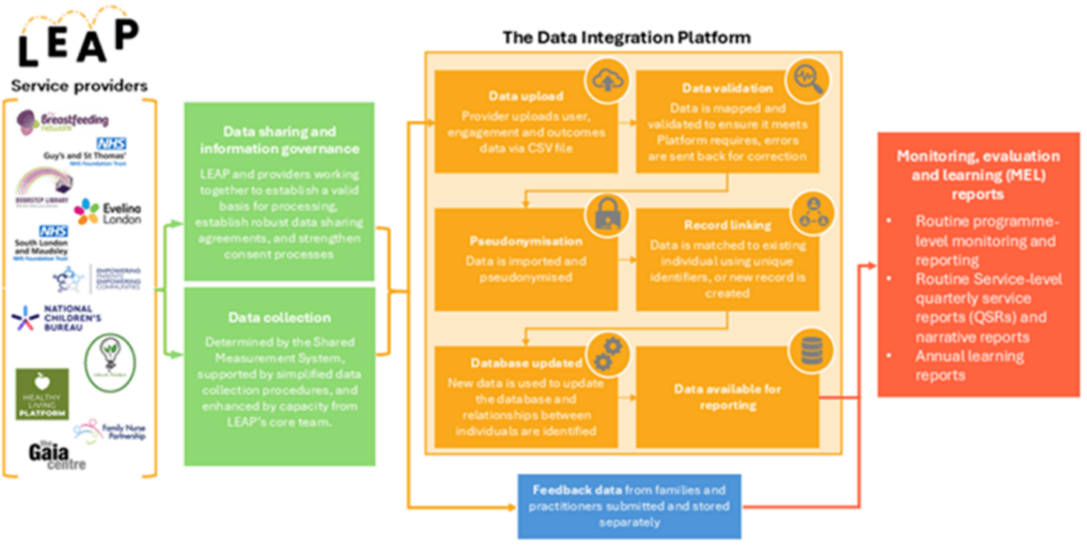
The Platform was never intended to be an integrated case management system. Participants were not identifiable, so it could only be used as a secondary data analysis tool. This project was not about identifying specific families. It was about reach, analysis and responding to specific evaluation questions.

# 4.0 Overview of Data Integration Platform and supporting functions

The next three sections of this learning journey set out three interrelated components that enabled the Data Integration Platform to function. As shown in Figure 2, they are:

1. The Platform’s technical components and approach to integration
2. A robust data-sharing and information-governance environment
3. A data-collection environment informed by the Shared Measurement System

Section 8 then describes the learning made possible by The Platform (set out in red in Figure 2).



**Data sharing and information governance**
LEAP and providers working together to establish a valid basis for processing, establish robust data sharing agreements, and strengthen consent processes

**Data collection**
Determined by the Shared Measurement System, supported by simplified data collection procedures, and enhanced by capacity from LEAP's core team.

**Data upload**
Provider uploads user, engagement and outcomes data via CSV file

**Data validation**
Data is mapped and validated to ensure it meets Platform requires, errors are sent back for correction

**Pseudonymisation**
Data is imported and pseudonymised

**Record linking**
Data is matched to existing individual using unique identifiers, or new record is created

**Database updated**
New data is used to update the database and relationships between individuals are identified

**Data available for reporting**

**Feedback data** from families and practitioners submitted and stored separately

**Monitoring, evaluation and learning (MEL) reports**
- Routine programme-level monitoring and reporting
- Routine Service-level quarterly service reports (QSRs) and narrative reports
- Annual learning reports

**Figure 2 – Overview of the Data Integration Platform and supporting components**

# 5.0 The Data Integration Platform: Technical components and approach to integration

## 5.1 Building The Platform

In 2019, LEAP invited tenders to build and host The Platform, eventually awarding the project to Lambeth Council. LEAP felt that it would benefit from Lambeth's resources and expertise, and also hoped that this decision would increase the likelihood of The Platform being sustained locally after the end of LEAP in 2025. This arrangement also provided early reassurance from an Information Governance (IG) perspective. Many services within LEAP's portfolio were already being delivered by Lambeth Council, and therefore possessed well-established data sharing processes.

Throughout The Platform's development, the IT consultancy, Fotheringham Associates, provided additional support and expertise. This was instrumental to the long-term success of The Platform.

As with the development of the Shared Measurement System (SMS) (https://leaplambeth.org.uk/reports/shared-measurement-system-learning-journey/), The Platform's development relied on extensive consultation with providers and local system leaders. This ensured that it built on existing practice and wasn't overly burdensome to professionals.

The Platform's development ran from 2018 to 2021, following the stages set out in Figure 3.

**Figure 3 – Overview of The Data Integration Platform's development**

## 5.2 Overview of data processing

Data moved through The Platform in the following steps. A visual representation of this process is available in Appendix 4.

1. **Data upload**: Service providers upload user, engagement and outcomes data
2. **Data validation**: Data is mapped and validated to ensure it meets platform requirements (errors are sent back for correction then resubmission)
3. **Pseudonymisation**: Data is imported and pseudonymised
4. **Record matching**: Pseudonymised data is matched to existing individuals using unique identifiers (new records are created where no match is identified)
5. **Database updated**: New data is used to update the database and relationships between individuals are identified
6. **Data available for reporting**

The following sections highlight some of the most important and innovative elements of this approach.

## 5.3 The Platform’s approach to linking diverse datasets which could not share identifiable data

The main challenge The Platform needed to overcome was being able to link diverse datasets that could not share identifiable data. As data was collected on the basis that it would not be shared with other organisations, LEAP was unable to onboard or analyse any data that included Personally Identifiable Information (PII).

Getting this right was crucial to LEAP’s ambitions for its routine Monitoring, Evaluation and Learning. Success would allow LEAP to build a programme-level view of families’ experiences of engaging with the programme, and to link individuals to long-term outcome data available within administrative datasets.

### A pseudonymisation approach

A pseudonymisation approach was considered to be the best option. This resolved the absence of an existing common unique identifier and the restrictions on using Personal Identifiable Information (PII) such as name, email address or date of birth.

Pseudonymisation is a technique which involves swapping identifiable data for non-identifiable data via an algorithm. It produces consistent results even from different data sources.

LEAP decided that the following datapoints would be used to generate a unique pseudonymised ID for use on The Platform:

- **For children**: Parent email address, child date of birth, child gender, and part of their first name[27]
- **For adults**: Their email address, NHS number and mobile number

The pseudonymised ID was created using the Secure Hash Algorithm 256 (SHA-256).

### Standardised pseudonymisation through the SHA-256 algorithm

Using the same pseudonymisation algorithm for all data, regardless of its source, was essential. If children’s or adults’ details were formatted in the same way, using the same algorithm would mean they would always be transferred into the same string of letters and numbers, allowing for consistent linking within The Platform.

A period of initial exploration identified the SHA-256 algorithm as an appropriate option. The SHA-256 algorithm is considered one of the most secure ‘hashing’ algorithms available. It transforms input data into a 256-character value that cannot be reverse-engineered. A ‘salt file’ (an additional random piece of data) was also used to further strengthen the pseudonymisation.

Importantly for LEAP’s purposes, the SHA-256 algorithm can be used natively via Structured Query Language (SQL) and on The Platform.

### Application of pseudonymisation

LEAP’s partners had varying levels of IT and data-system capacity. Accordingly, LEAP decided upon two approaches to pseudonymising data. These are described below and illustrated by Figure 4:

1. Pseudonymisation at source

This approach entailed information being pseudonymised prior to a partner sharing data with LEAP.

This approach was a requirement for obtaining IG approval for data sharing from NHS Trust partners. It was also used by Refuge (the national domestic abuse charity who ran LEAP’s Domestic Abuse Enhanced Caseworker service).

2. Pseudonymisation during the process of onboarding data to The Platform

In this approach, providers would share a data extract with LEAP staff, who would then apply pseudonymisation in a restricted data environment.

Due to limited capacity, this approach was used by all other services.

**Figure 4 – The Data Integration Platform’s two pseudonymisation approaches**

## 5.4 Uploading and validating data

Ensuring that data was of a high quality prior to being imported onto The Platform was crucial. It enabled higher link rates, reduced the probability of errors and allowed a greater volume of data to provide a more complete view of families’ experiences.

Providers were asked to **submit quarterly data extracts via a CSV file**. This is a plain, table-formatted, widely supported text file that imports and exports data between different software applications. This data would either already have been pseudonymised or would be pseudonymised later during upload to The Platform.

The LEAP team developed **data templates using Excel Power Queries** to then clean the data – making sure it was in the right format so it wouldn’t be rejected by The Platform.

CSV files were then **onboarded onto The Platform via Flatfile software**, which automatically mapped submitted data to expected data types and validated it – highlighting any anomalies, offering alternatives for selection or rejecting the data if it was of poor quality. This second layer of validation (on top of the Power Query templates) made sure that as little poor-quality data made it onto The Platform as possible – helping to avoid significant issues.

**Different data types went through more or less stringent validation**. Consent, gender and postcode,[28] for example, were crucial data fields without which onboarding could not proceed. Ensuring that data was of a high quality for the fields that pseudonymisation was based on was also vital, helping to ensure that there was a high match rate with data already in The Platform. Dates of birth, phone numbers and NHS numbers needed to always be in exactly the same format, so that they would produce the same pseudonymisation output. For example, ‘rogue’ spaces that would be perceived as text by the pseudonymisation algorithm needed to be removed.

In contrast, missing data in the minimum dataset (See appendix 3) was permitted. More complete datasets were expected to provide the missing characteristics when records were linked within The Platform.

Significant issues would be sent back to providers for correction and then resubmission.

## 5.5 Record linking

After data was uploaded to The Platform and pseudonymised (see above), a matching process was performed using the pseudonymised ID. This played out in the following way:

1. If a match was found with an existing record in The Platform (i.e. a record that shared the same pseudonymised ID), the existing parent or child record was updated with any new user data provided in the latest extract. This relied on a preset hierarchy to determine which data should be treated as most authoritative. NHS data was treated as the most reliable.
2. If no match was found, a unique LEAP individual reference was created, and a new individual was added to the dataset
3. Relationships with other individuals were identified
4. Any fields specific to the service in question are added (e.g. the latest outcomes data)

Later in The Platform's lifetime, LEAP introduced a Family ID. This linked children and parents with key shared details under an additional pseudonymisation code.

Family IDs were not, however, shared across all services. They were either service-specific, specific to an administrative dataset, or used by a subset of services (i.e. all Communication and Language Development services or maternity services covered by Badgernet).

This meant that a child attending a Speech and Language service in a childcare setting could be linked to their parent attending a Parenting Course. This allowed for a more accurate understanding of how an individual family had engaged with LEAP services.

## 5.6 Facilitating reporting: the 'Golden Dataset'

The Platform processed and stored data, but it was not interactive. Services, the funder and other interested stakeholders could not input or view data directly.

To facilitate routine reporting on Platform data, LEAP introduced the Golden Dataset. This was a single, central data model, published via Power BI, to which various routine reports linked in the form of connected tables. Examples included Quarterly Service Reports (QSRs) and programme-level reports (see Section 8).

Using a single central data model ensured consistency across reports by guaranteeing they were built on the same underlying source. The model also increased efficiency by having only one data model to maintain.

Non-routine evaluation activities (i.e. the collective impact analysis described in Section 8) did not rely on the Golden Dataset. They were run separately.

## 5.7 Reflections on the development of the Data Integration Platform

### Significant investment of time and resource

The Data Integration Platform took a long time to design, build, test and implement. It also demanded significant investment, including commissioning external consultants to support the development process. LEAP deemed that this investment was worthwhile given the potentially transformative impact of completing The Platform.

### Consistent quality assurance

Having Fotheringham Associates support The Platform's development process was instrumental. Members of the team provided consistent technical knowledge, support to resolve issues and advice on new approaches. Their knowledge and experience cannot be under-estimated as a contributor to the success of The Platform.

### Iteration to meet emerging requirements

The primary objectives for The Platform were clear from the start of the development phase and remained largely consistent throughout LEAP's lifetime. However, during the lifetime of the programme some additional requirements emerged organically over time. Theses required additional technical development cost. This occurred as the LEAP team became more familiar with the capabilities of The Platform and how it could support evidencing the programme Theory of Change.

Like all IT projects, software updates also caused some issues and additional work. The Flatfile software which provided intuitive data mapping and validation functionality as part of the uploading process, for example, underwent two major upgrades during the lifetime of The Platform, with a number of infrastructure changes also required. Fortunately, the foundations of The Platform were robust enough to meet emerging needs.

# 6.0 Developing a robust information governance and data sharing environment

The initial build and testing of the Data Integration Platform took around 12 months. Establishing robust information governance took much longer.

## 6.1 Key elements of information governance process

All providers who shared data with LEAP went through a rigorous information governance process. This ensured that all data processing was compliant with both organisational requirements, and UK General Data Protection Regulation (GDPR). This included:

### Establishing a valid legal basis for data processing

According to UK GDPR, all organisations must have a valid lawful basis for processing personal data.[29] Establishing this in the context of LEAP required lengthy and rigorous processes, especially for NHS trusts.

Most services shared data on the basis of consent. However, NHS trusts and other providers shared administrative data on the basis of legitimate interest – i.e. data was shared without explicit consent to fulfil a valuable purpose with minimal impact on privacy.

### Agreeing Data Sharing Agreements (DSAs) and Data Protection Impact Assessments (DPIAs)

To establish data sharing pathways, organisations typically need to establish a Data Sharing Agreement (DSA). When data processing involves higher levels of risk to privacy and personal freedoms, a Data Protection Impact Assessment may be required (DPIA). These documents record how information is being shared, and that the risks involved in doing so have been identified and mitigated.

DSAs were drafted for each provider whose data would be onboarded onto The Platform. In some cases, DPIAs were also carried out. These provided additional confidence in the security of the pseudonymisation software, which many providers had not encountered before.

Sharing data with external evaluation partners also led to updated agreements and additional DPIAs – including for LEAP's local evaluation partner Dartington Service Design Lab, and the national evaluation led by the National Centre for Social Research.[30]

Establishing these agreements was often extremely time consuming, demanding close attention to every element of data processing. The most difficult and time-consuming approvals to obtain were from NHS Trusts.

### Strengthening consent processes

To ensure that LEAP's use of data was compliant with relevant legislation, LEAP refreshed its service-level data processing and consent processes. To achieve this, in 2019, Fotheringham Associates ran two workshops with each service.

They mapped out existing data collection processes and identified and rectified any consent-related or other data-related weaknesses. They also guided services on what new forms and processes needed to be created and set up a range of templates to make this as easy as possible.

This was further strengthened by the introduction of standardised consent forms in 2019. Some sets of services – e.g. LEAP's Communication and Language Development services that worked in local early years settings – also introduced collective consent forms to reduce the burden on both families and practitioners.

## 6.2 Reflections on establishing a robust data sharing environment

### Pushing the boundaries to make data sharing acceptable

Establishing complex data sharing processes sits firmly outside ‘business as usual’ for most statutory or voluntary and community sector organisations. This issue is not unique to Lambeth, with many organisations lacking the capacity, confidence and culture to readily establish data sharing processes.[13]

In this context, LEAP needed to invest significant time and energy into building partners’ motivation and ability to share data. This involved building relationships with key individuals, as well as significant patience and recognising that every request had to be negotiated and justified against wider priorities.

Establishing data sharing processes and agreements for administrative datasets – as well as obtaining the actual data – proved most challenging. This was particularly true of health visiting and maternity data from NHS trusts. External factors added further layers of complexity. For example, the LEAP area fell within the catchment of two NHS trusts for maternity care both of which introduced a new IT system (see below).

In some cases, reminding partners of why it would benefit them to share data helped to accelerate progress. For example, the fact that the Parent and Infant Relationship Service would receive its own dedicated evaluation using Platform data helped to get data sharing agreements with South London and Maudsley NHS Trust over the line. The consistency of LEAP’s data team was also instrumental, allowing LEAP to develop trusting relationships with organisations over time.

### Mitigating concerns around sensitive data

Certain partners were more cautious than others about sharing data with LEAP. This was particularly the case for those who processed particularly sensitive data, including the GAIA Centre run by Refuge (host of LEAP’s Domestic Abuse Enhanced Caseworker service) and NHS trusts. Establishing agreements tended to take the longest with these partners. DSAs and DPIAs went through many iterations and required sign-off at the highest level to ensure compliance.

### Inflexible and slow-moving IT systems

It is widely recognised that the quality of IT systems, and the capacity of staff working on them, varies significantly across the public sector.[31]

LEAP encountered multiple issues with slow-moving IT departments. For example, it took some partners a significant amount of time to install the pseudonymisation software, despite this being a simple programme. Similarly, one of the most damaging delays to data sharing was the introduction of Epic. Epic is an electronic patient records system used across Guys and St Thomas’ and Kings College Hospital NHS Trusts. During its installation, all other information governance work – including sharing data with LEAP – was sidelined. This was a significant impediment to evaluation timelines, delaying the sharing of administrative data by over a year. At the time of writing, LEAP had still not received any data from one NHS trust despite having completed DSAs, DPIAs and met other key IT requirements.

### A partnership approach and access to senior figures

Many of the above barriers were resolved or expediated by LEAP’s partnership working approach. This meant that LEAP had access to a network of influential local leaders who were able to promote the concept of The Platform. They helped push data sharing agreements through their respective information governance teams.

For example, this approach helped to secure vital meetings with Caldicott Guardians for NHS partners and key IT teams for statutory partners. Caldicott Guardians are senior figures responsible for protecting the confidentiality of people’s health and care information. This led to something of a ripple effect, with senior approval from one organisation helping others to buy into The Platform.

# 7.0 Data collection and upload: how data made its way onto The Platform

The key goal for LEAP's Data Integration Platform was to link individual-level data from across LEAP's different services and administrative datasets to enable a programme-level view.

The Platform itself did not prescribe what data should be collected and processed at a service level, nor how this should be done.[32] Instead, this was achieved via LEAP's Shared Measurement System (SMS) (https://leaplambeth.org.uk/reports/shared-measurement-system-learning-journey/), as set out at a high level in Section 3.

This section explores how data, structured and organised by the SMS, was collected and made its way onto The Platform. More detail on this work, as well as its strengths and weaknesses, is available in Section 5 of the SMS Learning Journey (*Putting the Shared Measurement System into practice*).

## 7.1 The need to build capacity and change cultures around data collection

The success of both the SMS and The Platform depended on the consistent collection of significant quantities of high-quality data. For The Platform to function, data needed to be of certain quality to enable a high link rate. Additionally, the more data in The Platform, the more sophisticated LEAP's analysis could be.

However, at the time The Platform was first launched, there was significant variation in LEAP services' capacity and ability to collect high-quality data. This was particularly the case when it came to using sophisticated outcome measurement tools. Some services had never had to collect large amounts of consistent data on an ongoing basis. Others lacked the IT systems to process data. Many felt under significant pressure due to contextual factors like high staff turnover.

LEAP took several approaches to addressing this, beyond simply requiring services to collect data in formal task orders. This included involving services in deciding what data they would be collecting, offering ongoing encouragement and support, and making it as easy to collect and upload data. This is expanded on in the SMS Learning Journey (https://leaplambeth.org.uk/reports/shared-measurement-system-learning-journey/).

## 7.2 Simplifying and adapting data collection and upload processes

LEAP aimed to do everything it could to make accurate and consistent data collection as easy as possible. It introduced simple frameworks, tools and adaptable systems to upload data to The Platform. This included:

### Bespoke Monitoring, Evaluation and Learning (MEL) frameworks

Each service was provided with a bespoke MEL framework that was intended to act as a 'measurement toolkit', setting out how, when and from whom to collect data. This also included paper versions of data collection tools.

### Automated upload to The Platform via a digital Outcomes Assessment Tool

To support consistent data collection, LEAP commissioned Fotheringham Associates to develop a digital outcomes assessment tool (known as The App). For each service, this included data fields corresponding to each feature of services' measures.

Data could be inputted directly into The App at the moment of collection or inputted later from paper versions of measurement tools. Data from The App was then downloaded onto a CSV file by LEAP staff and uploaded onto The Platform on a quarterly basis (alongside CSV files including other data types, as explored in Section 5).

### Paper-based processes for services without digital capability

Some services were unable to use The App. This was the case if practitioners didn’t have easy access to a digital device while delivering their service. In these instances, outcomes data was collected using paper forms, which were then inputted into The App by LEAP staff.

The Speech and Language Therapy (Evelina Award) team, for example, collected all WellComm toolkit[33] outcomes data using paper forms. This happened because the service was delivered on site at busy early years settings.

## 7.3 Quality assurance, support and additional capacity from an expanded LEAP core team

Introducing the SMS and Platform represented a step change in what was being asked of services. It also increased the amount of data available for analysis. LEAP supported these changes by restructuring and expanding its data-focused teams.

As part of this, LEAP introduced two new roles to offer quality assurance of data collection, as well as hands-on support for services to collect and share high quality data. This included working over an extended period to build practitioners’ confidence around using measurement tools, troubleshooting any issues, and offering direct support with data collection.

# 8.0 Learning from LEAP's Data Integration Platform

The Data Integration Platform was fundamental to LEAP's approach to routine Monitoring, Evaluation and Learning throughout its second half.

The Platform successfully linked individual-level data across services and administrative datasets. In doing so, it enabled sophisticated and accurate insights into how different families engaged with and benefitted from LEAP services and activities.

Crucially, The Platform helped to bring LEAP's Shared Measurement System (SMS) (https://leaplambeth.org.uk/reports/shared-measurement-system-learning-journey/) and Theory of Change to life. This allowed LEAP to analyse data from unique beneficiaries across its six priority data types – as well as using linked data to show how LEAP services were working collectively to support families. Linking to administrative datasets also allowed for evaluation of LEAP's impact over a longer timescale.

## 8.1 Learning that it was possible to meaningfully integrate data for analysis and evaluation

By the end of the programme, 31 service datasets and 4 administrative datasets had been successfully onboarded onto The Platform. Linking individual-level data across these datasets provided a powerful tool for both analysis of the LEAP programme, as well as local services more generally.

The Platform also enabled the linking of health and education data, rectifying a perennial problem faced by many local systems.

## 8.2 Routine service- and programme-level monitoring

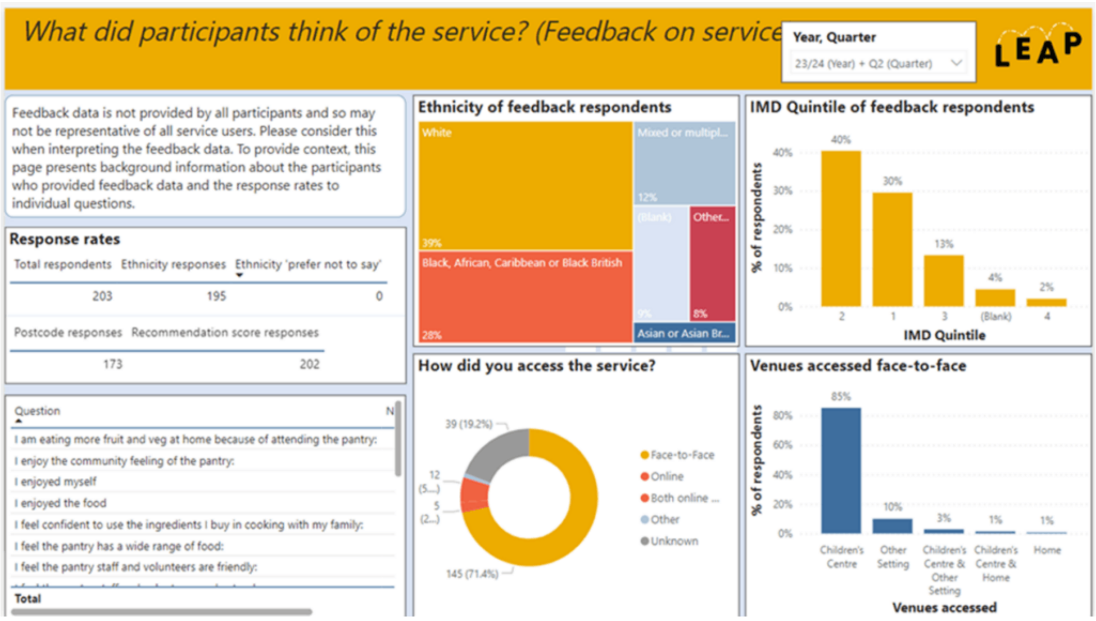
The Platform was crucial to LEAP's routine Monitoring, Evaluation and Learning approach at both the service and programme level. Prior to The Platform, LEAP's inability to link individual-level data either over time or across services meant there was often double counting across all data types.

For the purposes of routine monitoring, Platform data was fed into the Golden Dataset, which underpinned a range of neat summary reports at the service and programme-level. These reports used integrated data to ensure reporting included only unique beneficiaries. They also integrated data from administrative datasets to compare reach and user characteristics against local data.

### Programme-level monitoring to fulfil funder requirements

As set out in Section 3, one of the key challenges that The Platform was trying to address was LEAP's inability to accurately report on reach figures. This was a crucial requirement from the LEAP's funder, the National Lottery Community Fund.

Linking data across services resolved this, with the use of a unique identifier meaning individuals were only represented once in reach figures. Overall reach figures (including a summary of socio-demographic data) were reported to the funder alongside combined data on feedback, workforce, services and activities on a quarterly and annual basis. Figures 5 and 6 show some examples of programme-level reporting for the second quarter of 2023/24.



Year, Quarter: 23/24 (Year) + Q2 (Quarter)

Feedback data is not provided by all participants and so may not be representative of all service users. Please consider this when interpreting the feedback data. To provide context, this page presents background information about the participants who provided feedback data and the response rates to individual questions.

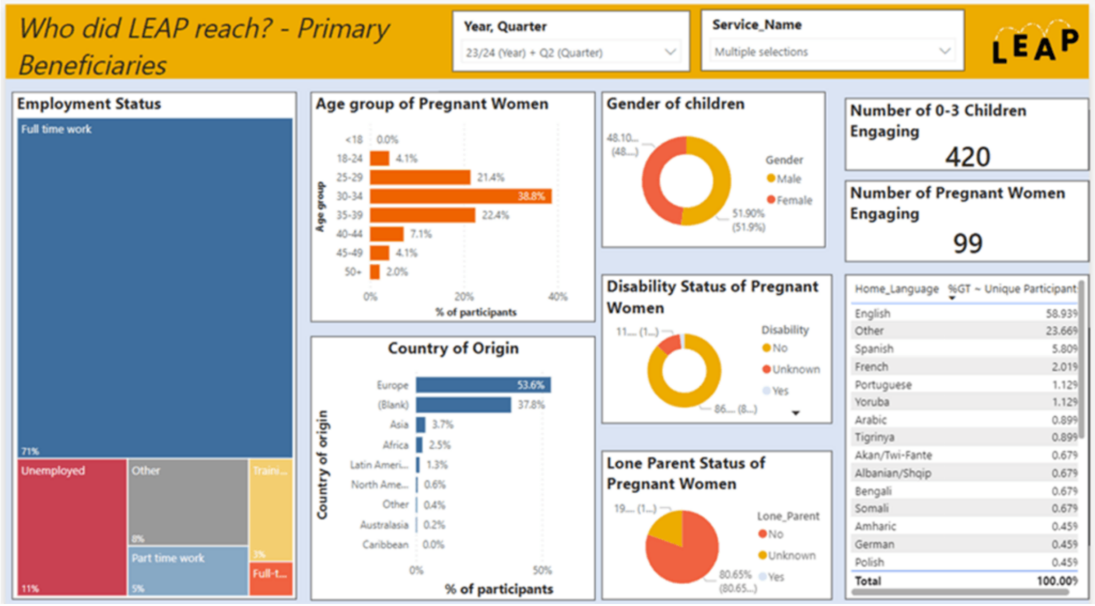
**Response rates**

| Total respondents | Ethnicity responses | Ethnicity 'prefer not to say' |
|---|---|---|
| 203 | 195 | 0 |

| Postcode responses | Recommendation score responses |
|---|---|
| 173 | 202 |

| Question | N |
|---|---|
| I am eating more fruit and veg at home because of attending the pantry: | |
| I enjoy the community feeling of the pantry: | |
| I enjoyed myself | |
| I enjoyed the food | |
| I feel confident to use the ingredients I buy in cooking with my family: | |
| I feel the pantry has a wide range of food: | |
| I feel the pantry staff and volunteers are friendly: | |
| **Total** | |

Figure 5 – Programme level report on feedback data, Q2 2023/24



Year, Quarter: 23/24 (Year) + Q2 (Quarter)

Service_Name: Multiple selections

**Number of 0-3 Children Engaging**
420

**Number of Pregnant Women Engaging**
99

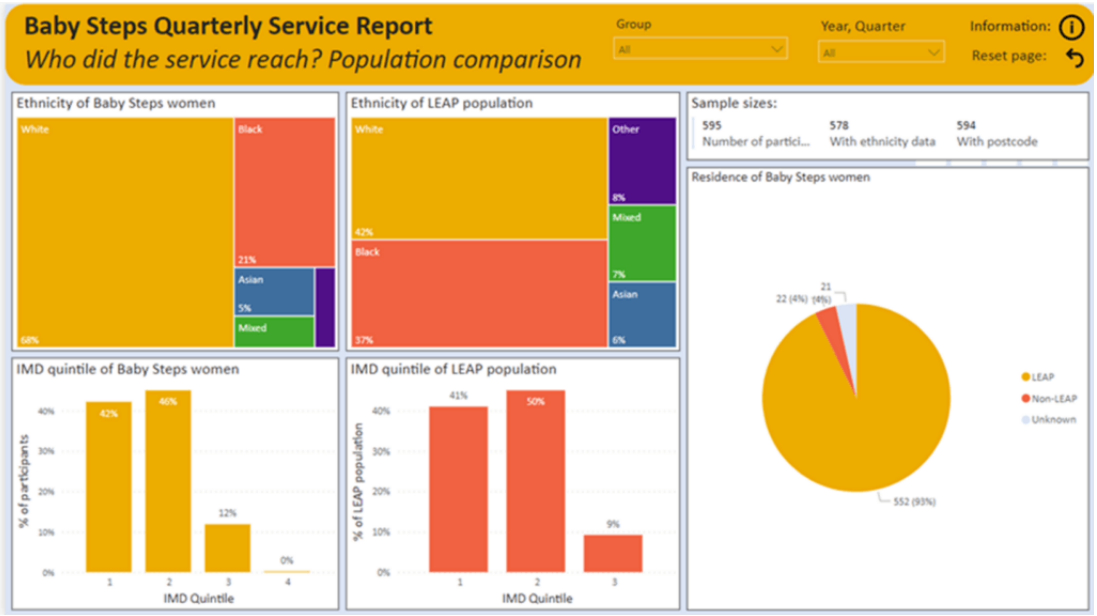
| Home_Language | %GT ~ Unique Participant |
|---|---|
| English | 58.93% |
| Other | 23.66% |
| Spanish | 5.80% |
| French | 2.01% |
| Portuguese | 1.12% |
| Yoruba | 1.12% |
| Arabic | 0.89% |
| Tigrinya | 0.89% |
| Akan/Twi-Fante | 0.67% |
| Albanian/Shqip | 0.67% |
| Bengali | 0.67% |
| Somali | 0.67% |
| Amharic | 0.45% |
| German | 0.45% |
| Polish | 0.45% |
| **Total** | **100.00%** |

Figure 6 – Programme level report on service activity, Q2 2023/24

Service-level monitoring for data-informed decision making and improvement

Platform data, via the Golden Dataset, also fed into service-specific Quarterly Service Reports (QSRs). These reported on lifetime data across LEAP's priority data types, as prescribed by services' bespoke Monitoring, Evaluation and Learning (MEL) frameworks.

The Platform ensured that this data included only unique beneficiaries, while linking of user records across services (as described in Section 5) ensured that key information on the characteristics of families and children was as accurate as possible. Figures 7 and 8 show excepts from QSRs for the Baby Steps service.



**Baby Steps Quarterly Service Report**

*Who did the service reach? Population comparison*

Group: All

Year, Quarter: All

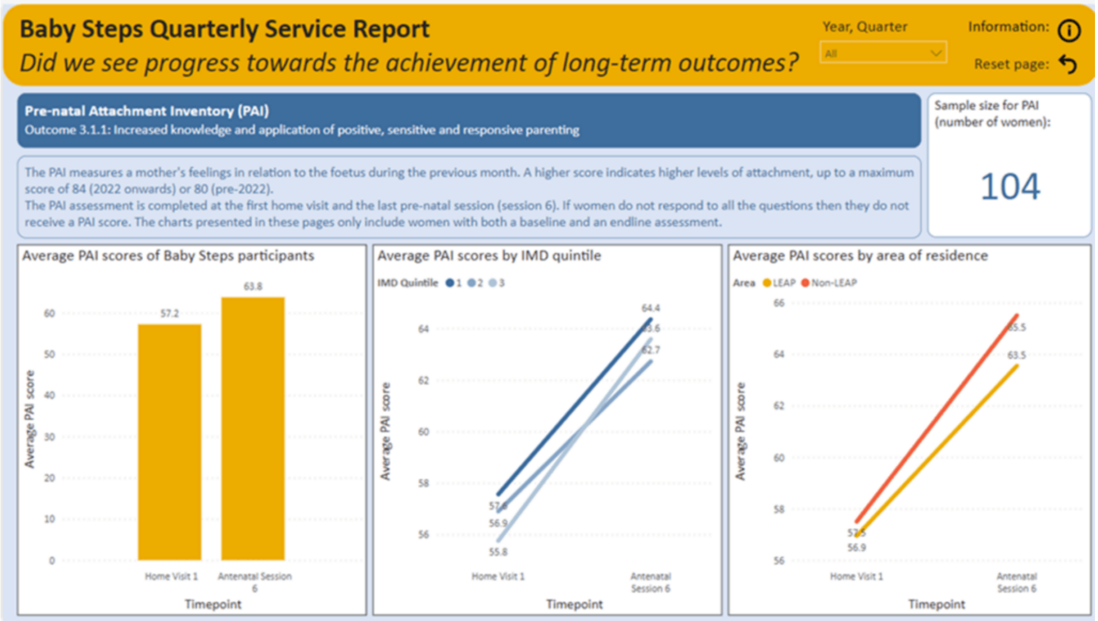
Sample sizes:

595 Number of partici...

578 With ethnicity data

594 With postcode

Figure 7 – Lifetime user data for the Baby Steps service, as seen through a QSR



**Baby Steps Quarterly Service Report**

*Did we see progress towards the achievement of long-term outcomes?*

Year, Quarter: All

**Pre-natal Attachment Inventory (PAI)**
Outcome 3.1.1: Increased knowledge and application of positive, sensitive and responsive parenting

The PAI measures a mother's feelings in relation to the foetus during the previous month. A higher score indicates higher levels of attachment, up to a maximum score of 84 (2022 onwards) or 80 (pre-2022).
The PAI assessment is completed at the first home visit and the last pre-natal session (session 6). If women do not respond to all the questions then they do not receive a PAI score. The charts presented in these pages only include women with both a baseline and an endline assessment.

Sample size for PAI (number of women): 104

Figure 8 – Lifetime outcomes data for Baby Steps (Pre-natal Attachment Inventory only)[34]

Services had access to their most recently updated data at any point. They were also encouraged to reflect on it via quarterly Service Data Meetings with the LEAP data team.

From a commissioning perspective, routine monitoring via QSRs allowed LEAP to easily hold services to account and suggest possible avenues for service improvement.

From a service perspective, having access to neatly summarised data allowed them to easily communicate their work to key internal and external stakeholders. They were also able to readily reflect on and improve their practice. Here are some examples scenarios:

- **User data indicates that services are failing to reach a target population**
  Staff deliberate on how they might reach families differently.

- **Progress towards long-term outcome measures varies across different groups of people**
  Services reflect on whether they could work differently with certain people.

- **Engagement data shows that few families are achieving dosage** (i.e. attending as many sessions as intended)
  Services reflect on how to rectify this.

## 8.3 Research, evaluation and programme-wide analytics

The Data linkage enabled by The Platform was also essential for LEAP’s wider portfolio of research, evaluation and programme-wide analysis.

This section focuses on LEAP’s two Annual Learning Reports. Both aimed to provide a programme-level view of LEAP’s activities, and assess whether its Theory of Change was working as intended

### 2021/2022 Annual Learning Report

LEAP’s first Annual Learning Report aimed to explore LEAP’s new SMS data, looking across services at a programme-level to:

1. assess the extent to which LEAP was reaching and engaging its target population;

2. summarise progress towards LEAP’s medium-term and long-term outcomes; and

3. identify areas in which the Theory of Change was not working as intended.

Data linkage enabled by The Platform not only ensured that individuals were only represented once in each piece of analysis, but also allowed LEAP to understand individuals’ journeys through the LEAP system. As a result, LEAP was able to address questions on the number, combination and sequencing of services accessed, and the effect this had on outcomes.

Key findings relating to reach and engagement included that:

- Families engaging with LEAP services broadly reflected the ethnic diversity of the population.

- The vast majority (85%) of children and families who engaged with LEAP attended only one service, with 14% attending between two and four services.

Example findings around outcomes within Domain 1 (Child Health and Development) included:

- The Parent and Infant Relationship One-to-One service was achieving an improvement in family relationships (assessed using the DC:0-5 Diagnostic Classification tool).[35]

- Children attending Natural Thinkers improved their wellbeing and involvement in activities between first and last engagement with the service (measured using the WellComm Toolkit).[36]

The 2021/22 Annual Learning Report is available online (https://leaplambeth.org.uk/reports/leap-annual-learning-report-2021-2022/).

## 2022/23 Annual Learning Report

LEAP's second Annual Learning Report's aimed to further decipher patterns of reach and engagement.

Linking data via The Platform was again crucial here. Analysts used individuals' unique IDs to link engagement data entries from different services and activities across the programme. This allowed them to see how individuals and families were accessing LEAP services. Example findings included:

- 67.2% of LEAP participants reached the minimum required number of sessions to have benefited from a service
- 22.5% of families accessed multiple services
- Families from LEAP's priority population were more likely to engage with multiple services than families not from the priority population

Analysts were also able to outline the most common combinations of services accessed by families who used two services. This is shown in Figure 9.

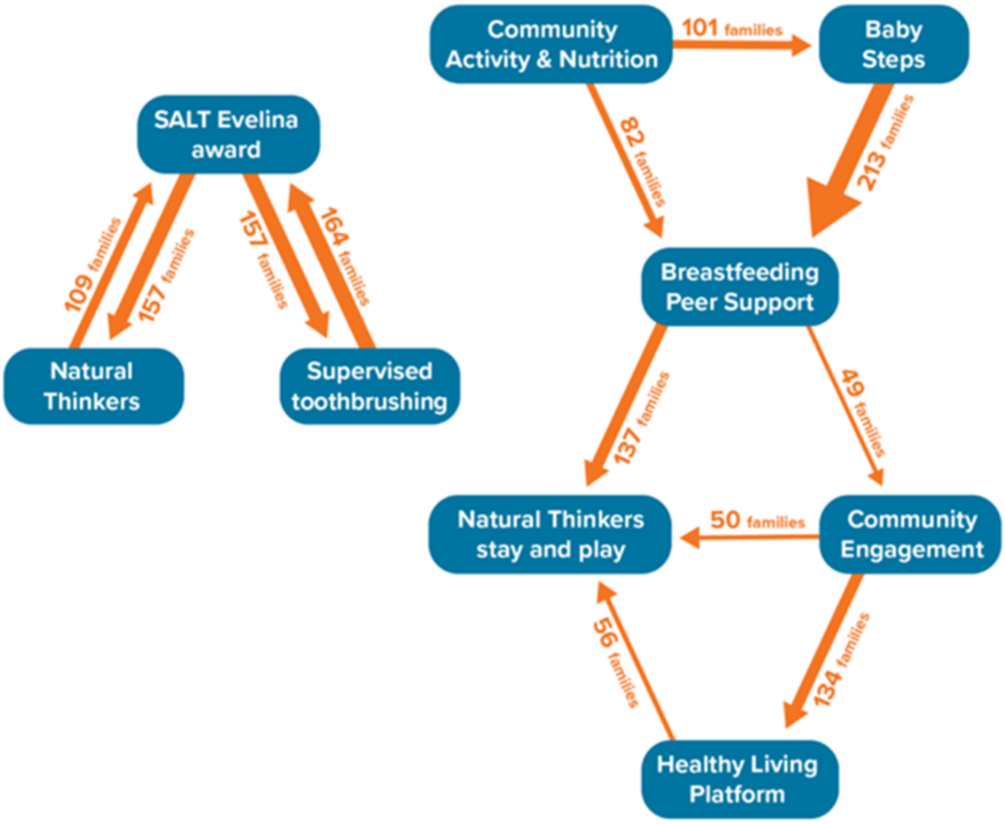


Figure 9 – Most common service use pathways for families who used multiple services

The 2022/23 Annual Learning Report is available online (https://leaplambeth.org.uk/reports/leap-annual-learning-report-2022-2023-landing-page/).

## 8.4 Collective impact analysis using linked administrative datasets

As part of LEAP's programmatic evaluation, the LEAP team analysed health visiting and early education outcomes for children engaged with LEAP services. They utilised the family ID linking capabilities of The Platform to connect individuals across service and local administrative datasets, assessing whether the extent of a family's engagement influenced their long-term outcomes.

The administrative datasets included in analysis were:

- ASQ-3 (a developmental screening tool used by Lambeth health visitors as part of a 2.5-year-old's health review)
- Early Years Foundation Stage Profile (EYFSP – a statutory assessment for children aged four and five)
- National Child Measurement Programme (NCMP – which measures the height and weight of children in reception and year 6).

Full analysis of NCMP data was halted due to a low match rate. This meant there were only 147 children from the LEAP area with a relevant LEAP engagement record.

Headline statistics included:

- LEAP children were 40% more likely to develop to expected levels than their peers.
- The more LEAP services a family engaged with, the likelier their two-and-a-half-year-old would reach at least expected levels of overall development.
- There were no significant differences in developmental outcomes for children at the end of reception compared to their peers. However, mitigating factors – including data limitations, greater levels of need, Covid-19 and the cost-of-living crisis – were likely to account for this.

The collective impact report is available online (https://leaplambeth.org.uk/reports/how-a-collective-impact-initiative-affected-the-outcomes-of-young-children-in-lambeth/).

## 8.5 Limitations to using Platform data

### Record linking still had a margin of error

The process of linking individuals across services and administrative datasets still had a margin of error. This was despite the overall strength of LEAP's pseudonymisation approach (particularly given the lack of a pre-existing unique child identifier).

In the context of the family ID, for example, it is possible that not all children who had a parent that engaged with LEAP were linked to that parent.

Matching to administrative datasets also incurred specific issues. The EYFSP dataset, for example, had a lack of identifiers common to LEAP's typical approach to generating a unique ID. This resulted in a lower rate of children linked between engagement data and the EYFSP dataset. As explored above, there was also a very low match rate for NCMP data, reducing the scope of possible analysis.

### Incomplete datasets

Service-specific data was only uploaded onto The Platform if individuals gave their consent. Though the numbers who did not consent were very low (e.g. only 10% of Baby Steps participants did not consent), it is possible that this may have introduced bias into the data available on The Platform.

As explained in Section 5, if the key data fields required for pseudonymisation were missing, data would be rejected prior to upload to The Platform. This also reduced the representative of the sample of families on The Platform.

### Limitations to what can be understood from administrative datasets

For administrative data specifically, LEAP was unable to definitively claim that it was services' activities that had led to changes in outcomes.

The administrative datasets that LEAP monitored were collected at only one time point for each child. To be able to make robust causal claims, LEAP would have needed to measure changes in the same outcome before and after engagement with LEAP. It would also have needed to compare these changes to a control group (ideally randomised). Without this, LEAP was only able to speak of 'associations' between service engagement and developmental outcomes.

This limitation was clearest when grappling with the finding from the collective impact analysis (see above). The analysis found that children who engaged in targeted services, or social and emotional development (SED) services, had poorer outcomes compared to their peers who had not engaged with LEAP services.

Without knowing children's starting point before their engagement with LEAP, it was impossible to determine whether 'poorer outcomes' still represented an improvement compared to families in similar circumstances who had not engaged with LEAP.

## 8.6 Opportunities for further learning

Given resource and time constraints, LEAP was unable to make use of the capabilities of The Platform in all the ways they had hoped to.

Work that LEAP was unable to conduct included linking engagement data to other population-level datasets such as Health Visiting Carenotes data and Children's Centre records. This could have included data that offered more insight into wider social determinants of health, including housing and parental employment.

LEAP was also unable to explore whether engagement with certain groups (or pathways) of services impacted long-term outcomes.

# 9.0 Sustainability and influencing

Securing a long-term future for the Data Integration Platform was a key aim since its conception in 2018. LEAP also hoped to use the success of The Platform to conduct national influencing work, helping other organisations to use data to give children the best start in life.

## 9.1 Exploring opportunities for local sustainability

Senior LEAP staff explored several opportunities to secure the long-term local future of The Platform. This included long-term discussion with relevant teams at Lambeth Council and Lambeth Together (the local integrated care partnership).

Despite enthusiasm about the capabilities of The Platform, unfortunately they were unable to move forward in either case. A few key barriers were particularly influential, including:

- wider ongoing integration work putting a pause on all similar projects (as explored in Section 6);
- misalignment between key local priorities and The Platform's areas of emphasis (particularly its focus on outcomes, and on linking health and education data); and
- IT changes reducing the compatibility of The Platform with internal systems.

Despite this, lessons from The Platform's development did inform local integration activities. Lambeth Council, for example, adopted a similar approach to generating unique identifiers to improve data linkage across their databases.

## 9.2 National influencing

There has been significant appetite to learn from LEAP's Shared Measurement System (SMS) (https://leaplambeth.org.uk/reports/shared-measurement-system-learning-journey/) and Data Integration Platform at a national level. In 2022 the Data Integration Platform was spotlighted by Nesta as 'leading the way on data integration.'[12]

Stakeholders have expressed particular interest in learning from how LEAP successfully linked data without a pre-existing unique identifier. Many have also been impressed with the coherence of LEAP's SMS, particularly how LEAP managed to identify robust, and in some cases shared, long-term outcome measures.

LEAP staff have presented on The Platform at multiple national webinars. Learning from The Platform continues to inform the National Children's Bureau's policy work about multi-agency information sharing and the introduction of a Single Unique Identifier.

# 10.0 Key messages and reflections

**It worked!**

LEAP was able to successfully integrate 35 service and administrative datasets, enabling individual-level data to be linked across the LEAP programme and beyond. Through an innovative pseudonymisation approach and extensive work to build robust data sharing and collection environments, LEAP managed to overcome challenges that have long troubled the early years sector. This is a remarkable achievement, and has been rightly recognised as an example to learn from by national stakeholders.[10]

## Developing and implementing an innovative pseudonymisation approach to allow data linkage

Being able to successfully link diverse datasets that could not share identifiable data was The Platform's most crucial challenge. This is something that numerous organisations have struggled with, and there is no widely used method for doing so.

LEAP's pseudonymisation approach – using key identifiers shared across services and standardised pseudonymisation via the SHA-256 algorithm – was able to achieve this. As well as developing a robust pseudonymisation process, LEAP also iterated and adapted to make sure this approach could be used by a wide variety of services. This included reshaping The Platform architecture to allow for both at-source pseudonymisation and pseudonymisation during the process of onboarding data to The Platform.

## Transforming what was possible within LEAP's Monitoring, Evaluation and Learning

Before the Data Integration Platform, LEAP was unable to accurately report on its overall reach, fully appreciate its collective impact, or make use of the administrative datasets it had access to. This was due to the programme's inability to link individual-level data across datasets.

The Platform successfully addressed these challenges. Alongside the Shared Measurement System (SMS) (https://leaplambeth.org.uk/reports/shared-measurement-system-learning-journey/), it allowed for more accurate monitoring at both a service and programme level. It also enabled in-depth analysis to understand how LEAP services work together to produce long-term change for children and families. Consequently, LEAP was able to continuously improve its services, and the programme was able to share more sophisticated knowledge of 'what works' to support children's development.

## A Transferrable System

LEAP focused on the Early Years, but The Platform's capabilities do not have to be limited to this area. It can accommodate any dataset so long as key identiers are included. This provides opportunities for any combination of services so long as the same personal identifable information is used as part of the pseudonymisation process.

## Building a robust data environment, not just a platform

The Data Integration Platform's success relied on far more than just its technical components. LEAP also invested heavily in other enabling factors. It ensured that information governance was robust and acceptable to all partners (no mean feat in the case of NHS trusts), and introduced systems to ensure that the data being fed into The Platform was of a sufficient quality, quantity and consistency. This was an ongoing process throughout the second half of LEAP, reaffirming that data integration is an enduring endeavour, requiring long-term consistency and resource commitment from dedicated stakeholders.

## Changing cultures around data collection and sharing

Both data sharing and data collection did not come naturally to many providers. Many had limited skills, capacity and motivation to engage with The Platform, or deal with conflicting priorities. To overcome this, LEAP invested in building capacity and changing cultures within organisations over an extended period. This included:

- introducing processes and offering support to make tasks as straightforward as possible;
- getting buy-in from senior leaders who could help push the concept of The Platform; and
- patiently negotiating other priorities.

## Time and money

Developing and implementing the Data Integration Platform was a huge piece of work. As with many IT projects, it cost more and took longer than expected. Though the initial build took around a year, adding additional functionalities and negotiating data sharing lasted throughout LEAP's second half.

There were also many unexpected delays, particularly around securing access to NHS datasets. Cost was not limited to building The Platform. Collecting and analysing data also required ongoing investment over a number of years. Fortunately, LEAP was a well-resourced, 10-year programme, so was able to manage this.

The high levels of investment required to establish and conduct successful data integration are likely to be one of the key reasons why LEAP struggled to secure long-term local sustainability for The Platform, and will not be unique to the Lambeth context.

## Evidence of lack of appetite for integrated outcomes data

Despite the growing evidence of the value of data integration in the early years, LEAP had little success finding a long-term host for The Platform. A key reason for this was a lack of alignment between local priorities and the functionalities offered by The Platform. This included stakeholder emphasis on understanding service performance rather than linking engagement and outcomes data, and limited appetite for integrating health and education data.

# 11.0 Appendices

## 11.1 Appendix 1: Glossary of key terms

| | |
|---|---|
| Collective impact | A framework and approach whereby people and organisations work together and share information to address complex social problems and achieve a common goal. |
| Data Integration | A type of data sharing that involves record linkage, which refers to the joining or merging of data based on common data fields. |
| Indicator | Any quantitative or qualitative variable that provides a simple and reliable means to assess performance or achievement. Unlike a measure, this does not necessarily need to be a direct observation of a phenomenon. In the LEAP context, indicators primarily refer to quantitative assessments of LEAP's four performance data types (inputs, user, engagement feedback). |
| Measure | A direct observation or calculation of a particular attribute or phenomenon. In the LEAP context, measures were used in the context of outcomes for children and families. |
| Measurement tool | A measurement technique used to capture either an indicator or measure. |
| Place-based, collective impact: | Place-based collective impact combines a place-based focus on a geographic location, community engagement and local decision-making, with collective impact's emphasis on cross-sector collaboration, adaptive management and systems change. |
| Routine monitoring, evaluation and learning (MEL) | The routine collection of information to track progress on predetermined, essential indicators across six key data types (input, user, engagement, feedback, outcome and impact) at both programme- and service-level. |
| Shared Measurement | Shared measurement involves organisations working on similar issues, and towards similar goals, reaching a common understanding of what to measure, and collaboratively developing the tools to do so. |
| Theory of change | A process for thinking about an organisation or project's 'story,' logically linking target group, activities, outcomes and impact. It encourages organisations to consider how change happens in the short, medium and long term to achieve the intended impact. |

## 11.2 Appendix 2: LEAP's Theory of Change

LEAP's theory of change included the following components:

1. **An overarching programme-level Theory of Change**
   Describing LEAP's intended approach at the highest level. As shown in Figure 10, this included eight interconnected outcome domains that underpinned the programme's ultimate intended impacts.

2. **Eight domain-level Theories of Change**
   Breaking down each outcome domain into a range of long-term outcomes, and describing how LEAP's services, activities, and ways of working intended to contribute to them. The long-term outcomes within each domain are set out in Figure 11.

3. **20+ service-level theories of change**
   Outlining in more detail how service-specific activities fed into long-term outcomes.

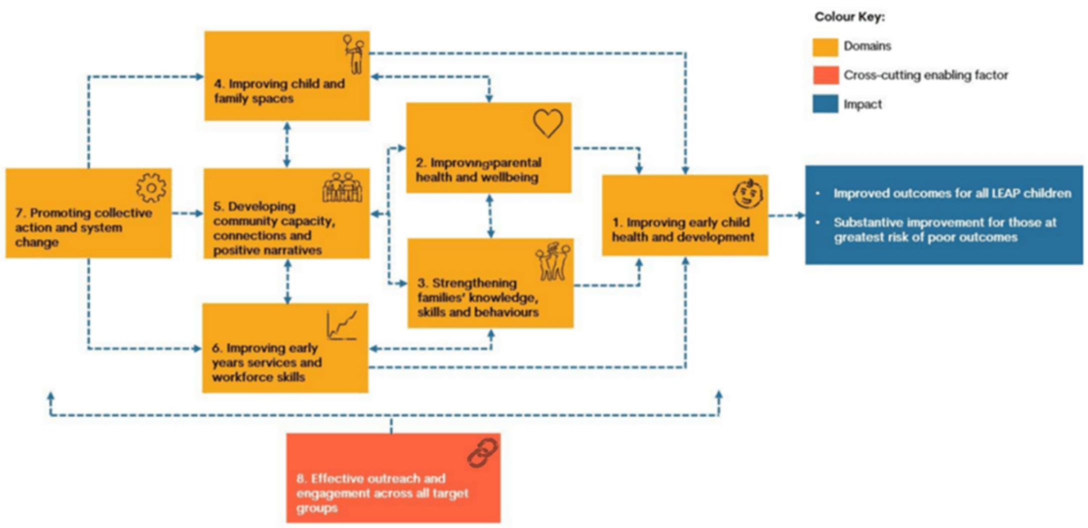
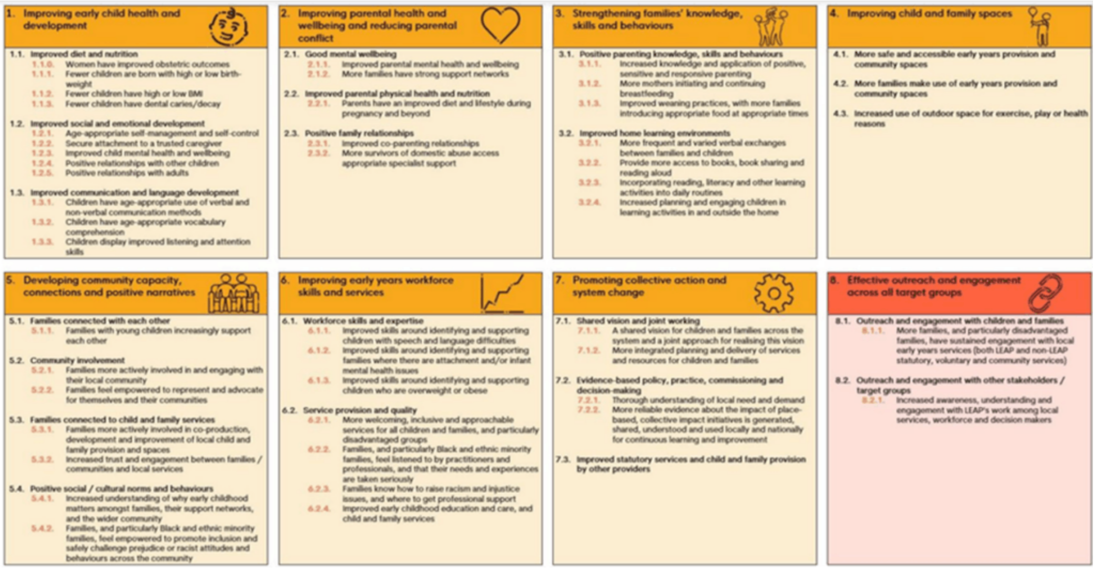


**Figure 10 – LEAP's programme-level Theory of Change**



**1. Improving early child health and development**

- 1.1. **Improved diet and nutrition**
  - 1.1.0. Women have improved obstetric outcomes
  - 1.1.1. Fewer children are born with high or low birth-weight
  - 1.1.2. Fewer children have high or low BMI
  - 1.1.3. Fewer children have dental caries/decay
- 1.2. **Improved social and emotional development**
  - 1.2.1. Age-appropriate self-management and self-control
  - 1.2.2. Secure attachment to a trusted caregiver
  - 1.2.3. Improved child mental health and wellbeing
  - 1.2.4. Positive relationships with other children
  - 1.2.5. Positive relationships with adults
- 1.3. **Improved communication and language development**
  - 1.3.1. Children have age-appropriate use of verbal and non-verbal communication methods
  - 1.3.2. Children have age-appropriate vocabulary comprehension
  - 1.3.3. Children display improved listening and attention skills

**2. Improving parental health and wellbeing and reducing parental conflict**

- 2.1. **Good mental wellbeing**
  - 2.1.1. Improved parental mental health and wellbeing
  - 2.1.2. More families have strong support networks
- 2.2. **Improved parental physical health and nutrition**
  - 2.2.1. Parents have an improved diet and lifestyle during pregnancy and beyond
- 2.3. **Positive family relationships**
  - 2.3.1. Improved co-parenting relationships
  - 2.3.2. More survivors of domestic abuse access appropriate specialist support

**3. Strengthening families' knowledge, skills and behaviours**

- 3.1. **Positive parenting knowledge, skills and behaviours**
  - 3.1.1. Increased knowledge and application of positive, sensitive and responsive parenting
  - 3.1.2. More mothers initiating and continuing breastfeeding
  - 3.1.3. Improved weaning practices, with more families introducing appropriate food at appropriate times
- 3.2. **Improved home learning environments**
  - 3.2.1. More frequent and varied verbal exchanges between families and children
  - 3.2.2. Provide more access to books, book sharing and reading aloud
  - 3.2.3. Incorporating reading, literacy and other learning activities into daily routines
  - 3.2.4. Increased planning and engaging children in learning activities in and outside the home

**4. Improving child and family spaces**

- 4.1. **More safe and accessible early years provision and community spaces**
- 4.2. **More families make use of early years provision and community spaces**
- 4.3. **Increased use of outdoor space for exercise, play or health reasons**

**5. Developing community capacity, connections and positive narratives**

- 5.1. **Families connected with each other**
  - 5.1.1. Families with young children increasingly support each other
- 5.2. **Community involvement**
  - 5.2.1. Families more actively involved in and engaging with their local community
  - 5.2.2. Families feel empowered to represent and advocate for themselves and their communities
- 5.3. **Families connected to child and family services**
  - 5.3.1. Families more actively involved in co-production, development and improvement of local child and family provision and spaces
  - 5.3.2. Increased trust and engagement between families / communities and local services
- 5.4. **Positive social / cultural norms and behaviours**
  - 5.4.1. Increased understanding of why early childhood matters amongst families, their support networks, and the wider community
  - 5.4.2. Families, and particularly Black and ethnic minority families, feel empowered to promote inclusion and safely challenge prejudice or racist attitudes and behaviours across the community

**6. Improving early years workforce skills and services**

- 6.1. **Workforce skills and expertise**
  - 6.1.1. Improved skills around identifying and supporting children with speech and language difficulties
  - 6.1.2. Improved skills around identifying and supporting families where there are attachment and/or infant mental health issues
  - 6.1.3. Improved skills around identifying and supporting children who are overweight or obese
- 6.2. **Service provision and quality**
  - 6.2.1. More welcoming, inclusive and approachable services for all children and families, and particularly disadvantaged groups
  - 6.2.2. Families, and particularly Black and ethnic minority families, feel listened to by practitioners and professionals, and that their needs and experiences are taken seriously
  - 6.2.3. Families know how to raise racism and injustice issues, and where to get professional support
  - 6.2.4. Improved early childhood education and care, and child and family services

**7. Promoting collective action and system change**

- 7.1. **Shared vision and joint working**
  - 7.1.1. A shared vision for children and families across the system and a joint approach for realising this vision
  - 7.1.2. More integrated planning and delivery of services and resources for children and families
- 7.2. **Evidence-based policy, practice, commissioning and decision-making**
  - 7.2.1. Thorough understanding of local need and demand
  - 7.2.2. More reliable evidence about the impact of place-based, collective impact initiatives is generated, shared, understood and used locally and nationally for continuous learning and improvement
- 7.3. **Improved statutory services and child and family provision by other providers**

**8. Effective outreach and engagement across all target groups**

- 8.1. **Outreach and engagement with children and families**
  - 8.1.1. More families, and particularly disadvantaged families, have sustained engagement with local early years services (both LEAP and non-LEAP statutory, voluntary and community services)
- 8.2. **Outreach and engagement with other stakeholders / target groups**
  - 8.2.1. Increased awareness, understanding and engagement with LEAP's work among local services, workforce and decision makers

**Figure 11 – LEAP's long-term outcomes by domain**

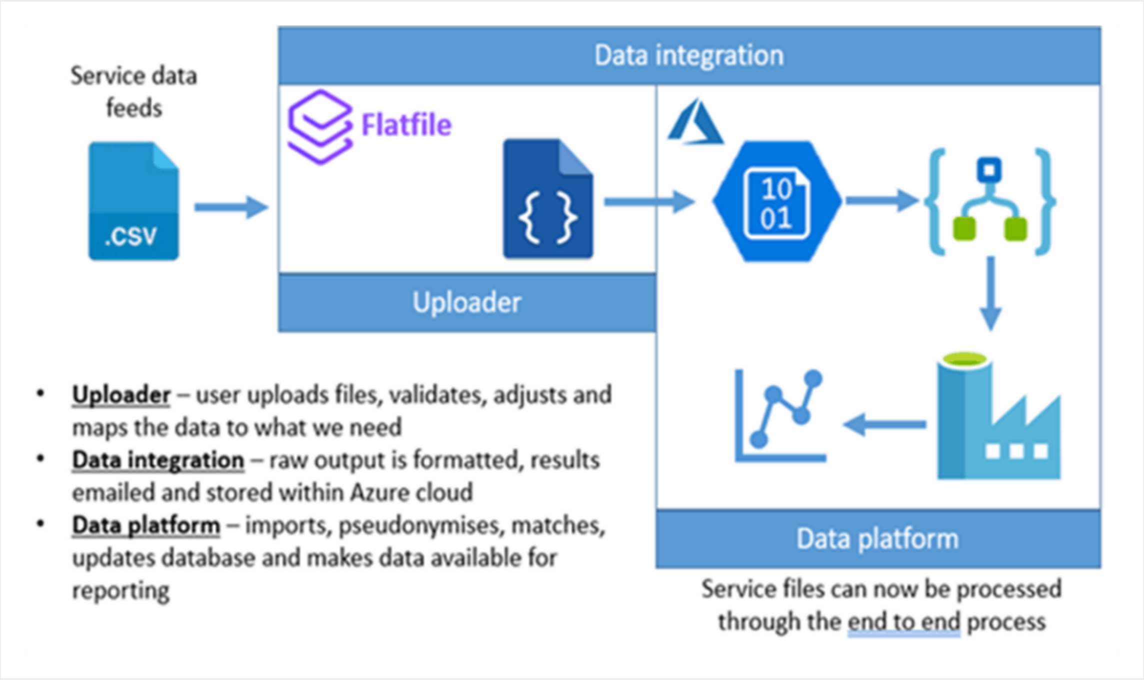
## 11.3 Appendix 3: LEAP's minimum dataset

LEAP's minimum dataset set out data which services needed to collect for all participants. Data types with an asterisk were collected by a subset of services.

| | |
|---|---|
| Adult Information | Adult email address<br>Role (Mother/Father/Primary Caregiver)<br>Main contact number<br>Additional contact number*<br>NHS Number*<br>Adult DOB<br>Gender<br>Postcode<br>Date registered with LEAP service<br>Volunteer status*<br>Lone parent status*<br>Employment status*<br>Country of origin<br>Home language<br>Disability status |

| | |
|---|---|
| | Ethnicity<br>Pregnancy status (and expected data due)<br>Parental education level*<br>Hours spent reading to child* |
| Child information | Primary carer e mail address<br>Child date of birth<br>Sex of child<br>NHS number*<br>Postcode<br>Data registered with LEAP service<br>Country of origin<br>Home language<br>Disability status<br>Ethnicity<br>Special Education Needs status* |

## 11.4 Appendix 4: Visualisation of Platform data processing



- **Uploader** – user uploads files, validates, adjusts and maps the data to what we need
- **Data integration** – raw output is formatted, results emailed and stored within Azure cloud
- **Data platform** – imports, pseudonymises, matches, updates database and makes data available for reporting

Service files can now be processed through the end to end process

# 12.0 References

① World Health Organisation. Global Strategy for Women's, Children's and Adolescents' Health; 2018. (Accessed 01 Feb 2025). Available from: https://iris.who.int/bitstream/handle/10665/276423/A71_19Rev1-en.pdf

↑ View

② National Lottery Community Fund. A Better Start through prevention and early intervention; 2024. (Accessed 01 Feb 2025). Available from: https://www.tnlcommunityfund.org.uk/media/insights/documents/A-Better-Start-through-prevention-early-intervention.pdf

↑ View

③ Cattan, S., Fitzsimons, E., Goodman, A., Phimister, A., Ploubidis, G.B. and Wertz, J. Early Childhood and Inequalities. Institute for Fiscal Studies; 2022. (Accessed 01 Feb 2025). Available from: https://ifs.org.uk/inequality/wp-content/uploads/2022/06/Early-childhood-inequalities-IFS-Deaton-Review-2.pdf

↑ View

④ Messenger, C. & Molloy, D. Getting it right for families: A review of integrated systems and promising practice in the early years. Early Intervention Foundation; 2014. (Accessed 01 Feb 2025). Available from: https://www.eif.org.uk/report/getting-it-right-for-families-a-review-of-integrated-systems-and-promising-practice-in-the-early-years

↑ View

⑤ Children's Commissioner. Utilising data to improve children's outcomes; 2022. (Accessed 01 Feb 2025). Available from: https://assets.childrenscommissioner.gov.uk/wpuploads/2022/12/cc-family-review-utilising-data-annex.pdf

↑ Instance 1 ↑ Instance 2 ↑ Instance 3 ↑ Instance 4

⑥ Actionable Intelligence for Social Policy. (2020). Introduction to Data Sharing and Integration. (Accessed 01 Feb 2025). Available from: https://aisp.upenn.edu/wp-content/uploads/2020/06/AISP-Intro-.pdf

↑ View

⑦ Royal College of Paediatrics and Child Health. State of Child Health in the UK; 2020. (Accessed 01 Feb 2025). Available from: https://stateofchildhealth.rcpch.ac.uk/

↑ View

⑧ Nesta. Harnessing the power of data in the early years: co-designing an early years data tool in Birmingham; 2024. (Accessed 01 Feb 2025). Available from: https://www.nesta.org.uk/report/harnessing-power-data-early-years/

↑ Instance 1 ↑ Instance 2

⑨ Martin-Denham, S., Scott, N., Horridge, K. & Pinney, A. Putting together the data jigsaw. University of Sunderland; 2023. (Accessed 01 Feb 2025). Available from: https://sure.sunderland.ac.uk/id/eprint/16347/1/Data%20Linkage%20FOI%20LA%27s%20Policy%20Brief.pdf

↑ View

⑩ Ibid.

↑ Instance 1 ↑ Instance 2

⑪ Children and Young People’s Health Policy Influencing Group. Integrated Care Systems and the health needs of babies, children and young people; 2024. (Accessed 01 Feb 2025). Available from: https://www.ncb.org.uk/sites/default/files/uploads/attachments/HPIG%20ICS%20report%20-%2031.01.2024.pdf

↑ View

⑫ Nesta. Unlocking the potential of early years data; 2022. (Accessed 01 Feb 2025). Available from: https://www.nesta.org.uk/blog/unlocking-the-potential-of-early-years-data/

↑ Instance 1 ↑ Instance 2

⑬ Department for Education. Improving multi-agency information sharing; 2023. (Accessed 01 Feb 2025). Available from: https://assets.publishing.service.gov.uk/media/64d500285cac65000dc2dd91/Improving_multi-agency_information_sharing_2023.pdf

↑ Instance 1 ↑ Instance 2 ↑ Instance 3

⑭ Department for Education. Munro review of child protection: a child-centred system; 2011. (Accessed 01 Feb 2025). Available at: https://www.gov.uk/government/publications/munro-review-of-child-protection-final-report-a-child-centred-system

↑ View

⑮ Children’s Commissioner. Connecting the dots – the importance of sharing data on children and families; 2023. (Accessed 01 Feb 2025). Available from: https://www.childrenscommissioner.gov.uk/blog/connecting-the-dots-the-importance-of-sharing-data-on-children-and-families/#_ftn1

↑ View

⑯ Health Equity North. An evidence-based approach to supporting children in the preschool years; 2024. (Accessed 01 Feb 2025). Available from: https://www.healthequitynorth.co.uk/app/uploads/CotN_Preschool_Report_9.pdf

↑ View

⑰ Academy of Medical Sciences. Prioritising early childhood to promote the nation’s health, wellbeing and prosperity; 2024. (Accessed 01 Feb 2025). Available from: https://acmedsci.ac.uk/file-download/16927511

↑ View

⑱ Department of Health and Social Care. Wider Determinants of Health; 2024. (Accessed 01 Feb 2025). Available from: https://fingertips.phe.org.uk/profile/wider-determinants

↑ View

⑲ Care Quality Commission. National review of maternity services in England 2022 to 2024; 2024. (Accessed 01 Feb 2025). Available from: https://www.cqc.org.uk/publications/maternity-services-2022-2024

↑ View

⑳ ECHILD. ECHILD; n.d. (Accessed 01 Feb 2025). Available from: https://www.echild.ac.uk/

↑ View

㉑ Department for Education. Improving multi-agency information sharing; 2023. (Accessed 01 Feb 2025). Available from: https://assets.publishing.service.gov.uk/media/64d500285cac65000dc2dd91/Improving_multi-agency_information_sharing_2023.pdf

↑ View

㉒ Policy in Practice. Multi Agency Safeguarding Tracker (MAST); n.d. (Accessed 01 Feb 2025). Available from: https://policyinpractice.co.uk/mast/

↑ View

㉓ National Children’s Bureau. Children and Young People’s Health Policy Influencing Group (HPIG); n.d. (Accessed 01 Feb 2025). Available from: https://www.ncb.org.uk/what-we-do/influencing-policy/policy-campaigns-and-projects/children-and-young-peoples-health#:~:text=In%20the%20past%20year%20we,what%20good%20practice%20looks%20like

↑ View

㉔ Department of Health of Social Care. Health and Care Act 2022; 2022. (Accessed 01 Feb 2025). Available from: https://www.legislation.gov.uk/ukpga/2022/31/contents

↑ View

㉕ Labour Party. Change: Labour Party Manifesto 2024; 2024. (Accessed 01 Feb 2025). Available from: https://labour.org.uk/wp-content/uploads/2024/06/Labour-Party-manifesto-2024.pdf

↑ View

㉖ UK Parliament. Children’s Wellbeing and Schools Bill; 2024. (Accessed 01 Feb 2025). Available from: https://bills.parliament.uk/bills/3909

↑ View

㉗ To account for twins.

↑ View

㉘ Postcode was deleted as the data is onboarded, being replaced with Lower Layer Super Output Areas (LOAS – a geographical unit compromising between 400 – 12000 households) to reduce any chances of identification.

↑ View

㉙ Information Commissioner’s Office. A guide to lawful basis; 2022. (Accessed 01 Feb 2025). Available from: https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/lawful-basis/a-guide-to-lawful-basis/

↑ View

㉚ National Centre for Social Research. A Better Start – National Evaluation; n.d. (Accessed 01 Feb 2025). Available from: https://natcen.ac.uk/ABS-national-evaluation

↑ View

㉛ National Audit Office Digital transformation in government: addressing the barriers to efficiency; 2023. (Accessed 01 Feb 2025). Available from: https://www.nao.org.uk/reports/digital-transformation-in-government-addressing-the-barriers/

↑ View

㉜ The design of the Platform did inform how LEAP services collected consent, and the user information necessary to generate a pseudonymised ID.

↑ View

㉝ The WellComm toolkit aims to assess whether the language children are using and understanding is appropriate for their age

↑ View

㉞ The Prenatal Attachment Inventory measures a mother’s feelings in relation to their foetus during the previous month. A higher score indicates higher levels of attachment.

↑ View

㉟ The Diagnostic Classification of Mental Health and Developmental Disorders of Infancy and Early Childhood (DC:0-5) is a measurement tool used to assess and diagnose mental health and developmental disorders in children from birth to five years old.

↑ View

㊱ The WellComm toolkit aims to assess whether the language children are using and understanding is appropriate for their age.

↑ View