LEAP Data Integration Platform Outline

Version: 1.0

*Contents Page*

[1 DATA PLATFORM CONTEXT 4](#_Toc12353564)

[1.1 Document purpose 4](#_Toc12353565)

[1.2 LEAP Overview 4](#_Toc12353566)

[1.3 The problem being solved 5](#_Toc12353567)

[1.4 Unique Beneficiaries 6](#_Toc12353568)

[2 SOLUTION 7](#_Toc12353569)

[*2.1* Solution Overview 7](#_Toc12353570)

[*2.2* Key features of the solution 7](#_Toc12353571)

[2.2.1 Pseudonymization 7](#_Toc12353572)

[2.2.2 Pseudonymization at source or destination 7](#_Toc12353573)

[2.2.3 Data quality is mandated 7](#_Toc12353574)

[2.2.4 Modern architecture 7](#_Toc12353575)

[3 FUNCTIONAL DESCRIPTION 8](#_Toc12353576)

[3.1 Processing 8](#_Toc12353577)

[3.1.1 Upload data 8](#_Toc12353578)

[3.1.2 Validate data 9](#_Toc12353579)

[3.1.3 Pseudonymise data 9](#_Toc12353580)

[3.1.4 Match individuals 10](#_Toc12353581)

[3.1.5 Data subject rights 11](#_Toc12353582)

[3.2 Reporting 13](#_Toc12353583)

[3.2.1 Management reporting 13](#_Toc12353584)

[3.2.2 Analytics 14](#_Toc12353585)

[3.2.3 Routine monitoring and evaluation reporting 14](#_Toc12353586)

[3.3 Platform management 14](#_Toc12353587)

[3.3.1 Security management 14](#_Toc12353588)

[3.3.2 Data storage 15](#_Toc12353589)

[3.3.3 Data management 16](#_Toc12353590)

[3.3.4 Platform support 16](#_Toc12353591)

[3.3.5 Platform development 16](#_Toc12353592)

[4 DATA MODEL 17](#_Toc12353593)

[4.1 Data model overview 17](#_Toc12353594)

[4.2 Draft data model entities 18](#_Toc12353595)

[4.3 Data validation and business rules 19](#_Toc12353596)

[4.3.1 Text handling 19](#_Toc12353597)

[4.3.2 Automatic updating of data items 19](#_Toc12353598)

[4.3.1 Standard value validation 19](#_Toc12353599)

[4.3.2 Date validation 19](#_Toc12353600)

[5 APPENDICES 21](#_Toc12353601)

[Appendix A - Data validation and business rules 21](#_Toc12353602)

[A.1 Adult entity 21](#_Toc12353603)

[A.2 Child entity 22](#_Toc12353604)

[Appendix B - Standard values for validation 24](#_Toc12353605)

# DATA PLATFORM CONTEXT

## Document purpose

The purpose of this document is to capture the discussions on the LEAP Data Integration Platform over the last six months.

## LEAP Overview

LEAP is a 10 year innovative programme, started in 2015, created to better the lives of thousands of children in the Lambeth community, focusing on four wards Stockwell, Coldharbour, Vassall and Tulse Hill.

Funded by the National Lottery Community Fund and working with partners locally and nationally, LEAP is a project that aims to support the social, emotional, communication and language development of babies and children, their diet and nutrition as well as parents’ wellbeing, their social networks and the strength of their communities and wider environment.

LEAP want to work with the community, giving them the control of what services they need and ways in which to improve life for their families.

At LEAP’s core is the community of Lambeth. As well as providing services to enhance early years care in the area, LEAP also works with local parents, carers, organisations, social workers and experts to make sure services are tailored for Lambeth’s needs. LEAP’s work is co-designed with regular focus groups and volunteering opportunities to ensure everyone gets a say on how services should be designed and run.

LEAP is hosted by the National Children’s Bureau and comprised of partnership organisations, including Lambeth Council, Clinical Commissioning Group (CCG), Public Health, King's Health Partners (KHP), as well as local voluntary organisations, community groups, parents, babies and children, schools, nurseries, statutory bodies and local police leaders.

The following diagram shows the many types of organisations that deliver LEAP services.



The LEAP interventions are delivered through a collection of services managed by one or more service providers. The services are collected into a series of themes as shown in the diagram below.



## The problem being solved

Currently, reporting systems for LEAP interventions are in silos. Individual providers send anonymised and aggregated data to LEAP on a quarterly basis. This data cannot be linked across LEAP’s services.

This creates a number of challenges:

* It prevents LEAP from building a full understanding of who accesses LEAP services and their patterns of engagement.
* It inhibits the ability to evaluate the impact of LEAP across the programme. Only the effect of individual services can be evaluated in isolation.
* Most critically, it does not enable accurate reporting on unique beneficiaries to the Funder

## Unique Beneficiaries

A key challenge within the project is defining an approach to uniquely identifying beneficiaries.

A number of methods to solve this issue have been considered including the use of standard identifiers such as names, postcodes, NHS numbers and mobile numbers. Each have their challenges and issues.

The current view to solving this challenge is:

* + For children = Using a key that consists of parent email address\*1, child dob and child gender\*2
  + For adults = Key is the Email address\*1

*\*1 – According to ONS 2018 figures 99% of age range 16-34 have an email address and email addresses have to be unique*

*\*2 - Same gender multiple birth children could cause an issue but the recorded numbers of these in the Lambeth borough according to the ONS are very small (<1%)*

# SOLUTION

## Solution Overview

The planned solution is to create a single platform that receives and combines data from the disparate services, matches individual beneficiaries across the services and makes available pseudonymised data to enable monitoring and evaluation.



## Key features of the solution

There are a number of key features of the solution that can be highlighted:

### Pseudonymization

As we will be using Personal Identifiable Information (PII) to form the unique identifiers this data has to be protected. We will use a pseudonymization technique to render the PII unrecognisable.

It should be noted that anonymisation techniques are also being considered as part of the overall solution and this will be explored during the development of the platform.

### Pseudonymization at source or destination

The platform will accept data files pseudonymised by the LEAP service provider before sending and it will pseudonymise data received in the clear by service providers.

### Data quality is mandated

The data files uploaded to the platform must pass agreed data validation checks. Any files that fail will be rejected and the service provider will need to rectify and re-submit the data.

### Modern architecture

It is expected that the data platform will be implemented within a modern IT architecture, throughout this document a Microsoft Azure Cloud has been assumed as that was felt to be the most likely infrastructure platform.

# FUNCTIONAL DESCRIPTION

This section seeks to describe the features of the LEAP data platform and the supporting service. The platform has been considered in three key areas shown in the graphic below.



## Processing

This is the main functionality of the data platform, it covers the full lifecycle from the Services logging into the platform to processing the received data and identifying and matching individuals.

### Upload data

The process starts with the Services logging into the platform and ends with the successful uploading of the data file, ready for validation.



* Step1 – The Service navigates to a URL that is provided to them
* Step 2 – They enter the username and password credentials that have been provided in order to access the data platform. The user will be sent a message to a mobile phone and the need to enter the code provided in the message.
* Step 3 – The user navigates to the file location and selects the file to upload and confirms.

If for some reason the file fails to upload due to connectivity issues (network or service outage) they will receive an error message.

### Validate data

This process carries out a series of validation checks on the file and on the rows to ensure it passes the data quality threshold.

As part of the process a copy of the data file is taken in order to help with troubleshooting if there are support issues. If data has not been pseudonymised at source there will be personal identifiable information (PII) present, but can only be accessed by a small number of data platform support staff.



* Step1 – The file submission will be checked against the last submission to ensure duplicate files are not submitted. The exact checks will be agreed during the development phase but could include checking file type, file size, number of rows and/or last amended date. File submissions will have a have specific filename format incorporating the service and quarter of the submission.
* Step 2 – Data validation checks are performed on all rows in line with data model requirements. There will be a standard set of data items and validations for the minimum data set and specific validation for service specific data. Each row will be validated as far as possible, the user will be informed if the file has passed validation. If there are errors these will be recorded and reported to the sender by email with a copy also submitted to a LEAP team mailbox.
* Step 3 – When the file has successfully passed the validation stage a copy of the file is taken and stored in a secure location. The copy of the file that was saved after the previous submission is permanently deleted.

### Pseudonymise data

This process uses the OpenP pseudonymisation component and creates scrambled versions of key personal information that will be used to identify individuals. A ‘salt file’ will also be used to further secure the pseudonymization, this salt file contains additional factor for the algorithm and will be made available securely to sites that require to pseudonymise at source.

If the file being received has been pseudonymised at source then this process will be skipped.



* Step 1 – There will be a range of PII fields affected by this process which will be annotated on the data model. Fields will be concatenated if necessary and the OpenP component will be called passing the relevant key field(s). The scrambled fields received as outputs will be added to the submitted file and the PII fields will be blanked out or removed.

### Match individuals

The received pseudonymised file will now be processed. Each row contains information about one or more individuals. This process identifies whether the individuals are already known or need to be added to the data platform. The data on the file is also used to capture relationships between the individuals.



* Step 1 – There maybe data relating to multiple individuals on each row of data. The individuals found on each row will be compared against the previous stored individuals. Each set of pseudonymised ‘keys’ (for example NHS number, mobile phone number or concatenated email address, child date of birth and gender) will be examined in turn.
* Step 2 – If the individual on the row is found its record is updated with additional information if any is available on the row. For example, if the row being processed has a mobile phone number and this is missing from the previously stored information it is added. Details of the fields that will be updated can be found in the Data Model appendix.
* Step 3 – As the individual has not been matched a unique reference is created which will act as the individual’s identifier within the data platform. The next available identifier is stored on a separate table and is incremented on creation of a new individual. Details of the unique reference can be found in the Data Model.
* Step 4 – After the unique reference has been generated the data pertaining to the individual is used to create that individual’s record within the data platform.
* Step 5 – Once the individuals have been processed for each row, the row is processed again to identify the relationships between the individuals. These are captured by writing records to the Relationships table, with entries added for each direction (e.g. a mother will be related to a child, and the child will be related to the mother). Some services may include individuals on multiple rows within the file, in these cases there will be data items available to help identify relationships.
* Step 6 – The row from the data file relates to a service as well as individuals and this step adds information to a table to record the involvement of the individuals in the service. After all records have been processed the service specific data on the file is also stored on the data platform in a service specific table.

### Data subject rights

The platform needs to manage the various rights of data subjects. Whilst it is the responsibility of the service providers to receive and initially process the requests from data subjects, the platform will need to receive and process these requests as a set of different transactions.

The table below illustrates the LEAP response to the rights of data subjects involved in the LEAP data platform:

|  |  |
| --- | --- |
| **Data Subject right** | **LEAP data platform response** |
| The right to be informed | Tailored consent forms, information sheet and privacy notice |
| The right of access | Covered in the Data Subject rights section |
| The right to rectification | Rectification is done within the service provider dataset and is processed by LEAP automatically on the next submission (inform Data Subject > 1 month) |
| The right to erasure | Covered in the Data Subject rights section |
| The right to restrict processing | Covered in the Data Subject rights section |
| The right to data portability | Covered in the Data Subject rights section |
| The right to object | Currently taking advice on applicability |
| Rights in relation to automated decision making and profiling | We are not carrying out this type of processing |

The general approach to satisfying the data subject requests will be to use the information provided to create a pseudonymised key using the standard data items to perform a look-up on the data platform. If information is found the request is processed, if no information is found then a ‘Data subject not found’ response will be returned. The data subject rights requests are submitted at a service level as they are made via a particular service provider.

Each data subject right that requires processing is discussed below.

#### The right of access

The data subject request results in the output of a report that details the information held about them. This will require a secure method of communication with the data subject that is still to be identified.



* Step 1 – The pseudonymised key is used to perform a look-up on the database
* Step 2 – If the record is found its information is added to a PDF report, note that this may include linking to a series of tables to access services and service specific data for the individual
* Step 3 – If no records are found then prepare a null report
* Step 4 – The request is completed by finalising and sending the report. Note: the approach to being able to send the report to the data subject has yet to be determined.

#### The right of erasure / withdrawal of consent

The data subject request results in the deletion of the pseudonymised information held for the data subject for the relevant service.



* Step 1 – The pseudonymised key is used to perform a look-up on the database
* Step 2 – If the record is found its information is deleted along with other linked information relating to the issuing of the request
* Step 3 – If no records are found then prepare a null report
* Step 4 – The request is completed by summarising a report that details the activity conducted

#### The right to restrict processing

The data subject request results in the records being flagged or unflagged as being restricted and set out of scope for analysis carried out on the data.



* Step 1 – The pseudonymised key is used to perform a look-up on the database
* Step 2 – If the record is found and the request is to add the restriction the restrict flag is set
* Step 3 – If the record is found and the request is to remove the restriction the restrict flag is removed
* Step 4 – If no records are found then prepare a null report
* Step 5 – The request is completed by summarising a report that details the activity conducted

#### The right to data portability

The data subject request results in the records being downloaded and made available in a simple CSV file format and sent to them. This will require a secure method of communication with the data subject that is still to be identified.



* Step 1 – The pseudonymised key is used to perform a look-up on the database
* Step 2 – If the record is found its information is added to a CSV extract file, note that this may include linking to a series of tables to access services and service specific data for the individual
* Step 3 – If no records are found then prepare a null report
* Step 4 – The request is completed by finalising the activity report and sending the CSV file. Note: the approach to being able to send the CSV file to the data subject has yet to be determined.

## Reporting

### Management reporting

There are a number of aspects to the management reporting that is required as part of the data platform, these should be rendered as a dashboard view with drill down into more detail if available.

#### Platform availability

This report shows the availability of the overall data platform, the upload processing and the reporting data.

#### Data uploads

This shows the attempted data file uploads identifying the Service, quarter and whether the file was successful. It should also be possible to see that multiple attempts were required in order to make the submission. If possible the errors for unsuccessful uploads should also be available for analysis on Service behaviour.

#### Processing summary

This shows a summary of successfully uploaded data files providing a view of the number of lines imported, the number of unique individuals added, the individuals matched with existing data and number of cases where pseudonymised information has been added (for example adding an NHS or mobile number).

### Analytics

LEAP data analysts will access the pseudonymised data on the platform by making a connection from analytic tools such as Power BI to the Azure database. No specific analytic reporting will be generated by the platform.

### Routine monitoring and evaluation reporting

The LEAP evaluation and monitoring team will access the pseudonymised data on the platform by making a connection from analytic tools such as Power BI to the Azure database and exporting specific queries at service, service cluster and programme level to enable analysis in a variety of software. This will allow them to verify service provider performance and alignment to their monitoring returns and evaluate programme impact and outcomes. No specific monitoring reporting will be generated by the platform.

## Platform management

### Security management

#### Two factor authentication

The data platform will operate with two factor authentication. The users will have a login and a password, but an additional factor, likely to be a mobile phone will be required to access the system. Users will typically be asked to enter a code sent to a mobile phone. By requiring the user to login using something they know (the password) and something they have (the mobile phone) it will increase the security of the platform.

#### Account management

Accounts will be allocated to approved individuals to access the platform. They will be issued with a user name and password.

There will be a self-service forgotten password facility where users can recover access to the platform.

It is expected that there will be at least four categories of users:

* System administrator – providing the ability to grant access to the system and control the operation of the platform
* Systems support – provides full access to data and processes running on the platform
* Platform contributor – provides access for the Services to upload data files and receive information on error messages
* Platform user – a LEAP user able to interrogate the pseudonymised data stored on the platform

#### Password policy

The password policy should include the following:

* Minimum password length (characters): 15
* Maximum password age (lifetime): Indefinite
* Passwords should contain 3 out of 4 of lowercase, uppercase, numbers, or symbols.
* Account should allow for 5 login attempts
* Account should be locked out for 30 mins

#### Audit

An approval process should be implemented before access is allowed to the area of the platform where datafiles held have personal identifiable information ‘in the clear’. This access should be automatically audited to ensure a record is kept of who had accessed this sensitive information. Notes should be kept against each session to record the detailed activity that the support staff has been undertaking.

Monitoring should be undertaken to check the approval records against the audit trails to ensure so the process is working as expected.

#### Security Monitoring

Logging that shows connections and usage of the platform should be kept and analysed to ensure that no unusual activity has taken place by unauthorised users. Any such activity should be investigated as a potential breach incident and LEAP should be informed. If this activity has potentially resulted in access of non-pseudonymised personal information then this should be flagged immediately with LEAP as they will need to notify their community of the breach.

### Data storage

The platform will have its own set of data tables to store pseudonymised data and supporting information. These tables will be made available to LEAP data analysts. The outline design for some of the key tables are provided in the Data Model section, although the final design will be determined during the build phase of the project.

A set of tables that record transactions into the platform will also be required to record operational information about files received including timestamp, size, number of rows file status etc.

The submitted files received from the service providers will be stored in original form until the next submission, at this time they will be deleted. They will be stored in a separate location and there will be no access to the files allowed for LEAP data analysts or anybody outside the support team and then only under strict audit requirements.

### Data management

The database underpinning the platform should be backed up or replicated to ensure that there is no loss of data. Any backed up or replicated data must be retained within a European data centre.

When the stored file submissions are to be deleted, on receipt of the next submission, the files must be permanently deleted and not merely logically deleted or references to them removed from the platform.

### Platform support

#### OpenP support

As the open source pseudomisation software, OpenP, is a critical part of the data platform the support and any ongoing maintenance of the software falls within the scope of the data platform support. This includes the SQL server plug-in, the desktop version and any upgrades required by future Windows versions.

#### Support resourcing

It is expected that dedicated support will be provided for the data platform. The support provided will include the range of skills that are essential for the effective operation of the platform. These will include database administration, cloud infrastructure support as well support of the specific functions of the platform.

#### Support categories

Support calls will be categorised in terms of urgency that will drive the expected operating level agreement for both acknowledgement and remediation of issues.

|  |  |  |
| --- | --- | --- |
| **Category** | **Impact** | **Required response** |
| Priority 1 | Issue that renders the whole data platform unusable | Fix within 8 hours |
| Priority 2 | Issue that renders part of the data platform unusable or degrades the service for some degree for all users | Fix within 24 hours |
| Priority 3 | Issue that degrades the service for certain users | Fix within 48 hours |

### Platform development

Resource will be required to build and commission the data platform in the first instance. The expectation is that the particular resources that take part in the development of the platform will form the core of the team that provide the support service for the platform.

Once the platform is operational the ongoing development that may be required due to new or revised requirements will be sourced from the dedicated support capacity. The exception to this approach will be if the developments required by LEAP are extensive and additional resourcing is agreed to focus on this need.

It is expected that the developments will follow an agile type methodology and work in open collaboration with LEAP and their partners.

# DATA MODEL

## Data model overview

Whilst the detailed data model will evolve during the development phase the conceptual level model has been drafted based upon conversations with the LEAP team over the last few months. The conceptual model shows the major entities within the platform and their relationships.



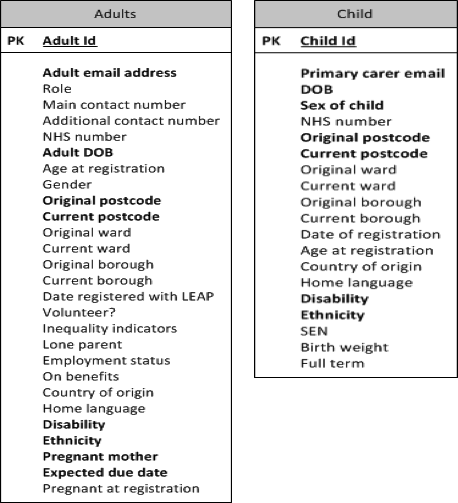
The table below provides a brief description of each table:

|  |  |
| --- | --- |
| **Entity** | **Description** |
| Child | Holds details of the children in the platform |
| Adult | Holds details of all adults discovered in the platform |
| Adult roles | Various roles that adults can play with respect to the children |
| Relationships | Record of the relationship between the child and an adult. This table will be downloaded and used if there is a need to attempt to identify family groups during evaluation. |
| Services | Holds the services or interventions that are part of LEAP |
| Service usage | This holds records of the usage of a service by adults and children. There will be records for each file submission. |
| Service detail | This entity will be specialized for each service, holding service specific information. There will be records for each file submission. |

Each of the tables in the conceptual model have a draft outline which are provided in the next section. It should be noted that the data model drafted here is focused on the input data as this has needed to be defined in order to prepare and amend service provider activities. It is likely that there will changes to the data model stored within the platform and available for analysis. As an example, in order to minimize personal data and lower the risk of identification of individuals if a breach occurred the postcode may be discarded after being used to derive the Lower Super Output Area.

## Draft data model entities

This section shows an outline of the main entities in the platform and the information that is expected to be captured and managed.





## Data validation and business rules

This section lists the typical validation that should take place on receipt of submitted file data for each data item. It also shows any business rules identified for that data item.

### Text handling

Prior to validating the data in each field, care must be taken to remove or ignore leading or trailing blanks from the data. No leading or trailing blanks should be added to the stored data platform database.

Note also that blanks may or may not appear in the middle of postcode fields, this should be handled appropriately to ensure validation can take place.

### Automatic updating of data items

When processing a subsequent submission or another service there maybe the opportunity to enrich the data held for individuals. Possible treatments for the key data entities are shown in Appendix A, but these are currently being reviewed by LEAP and will be confirmed later in the project.

### Standard value validation

Many code type data items have a set of standard values that will be used for validation prior to the uploaded file being processed. These data items and list values are shown in Appendix B.

### Date validation

In addition to checking that dates are in the correct format dd/mm/cccc and are a valid calendar date there is a check that the dates are reasonable for their context:

* Adults: should be between 01/01/1920 – 31/12/2009
* Children: should be between 01/01/1996 – 31/12/2026
* LEAP registration dates between 01/01/2015 – 31/12/2026

# APPENDICES

## Appendix A - Data validation and business rules

### A.1 Adult entity

| **Field** | **Validation** | **Business rules** |
| --- | --- | --- |
| Adult ID | Autogenerated | Generated by the platform. ID design to be decided. |
| Adult email address | Must be a valid email address layout |  |
| Role | Must be a valid value |  |
| Main contact number | Must be a valid mobile number layout | If missing and found on a matched record, update |
| Additional contact number | Must be a valid mobile number layout | If missing and found on a matched record, update |
| NHS Number | Must be a valid NHS number, 10 digits and conform to Modulus 11 algorithm | If missing and found on a matched record, update |
| Adult DOB | Must be a valid date format, a valid date and a reasonable date (see specific date note below) |  |
| Gender | List value: Male, Female |  |
| Original Postcode | Must be a valid postcode. | Always stored as indicator of a LEAP participant |
| Current postcode | Must be a valid postcode. | If different on a new matched record overwrite with new data |
| LEAP child? | Yes or No | Set when a LEAP postcode is first detected |
| Original ward | Must be a valid London borough ward | Derived from original postcode |
| Current ward | Must be a valid London borough ward | Derived from current postcode |
| Original borough | Must be a valid London borough | Derived from original postcode |
| Current borough | Must be a valid London borough | Derived from current postcode |
| Date registered with LEAP | Must be a valid date format, a valid date and a reasonable date (see specific date note below) |  |
| Volunteer? | Yes or No | If No and Yes found on a matched record, update |
| Inequality indicators | To be defined later in the project. Unknown allowed. | If unknown and found on a matched record, update |
| Lone parent | Yes or No. Unknown allowed | If unknown and found on a matched record, update |
| Employment status | Must be a valid value. Unknown allowed | If unknown and found on a matched record, update |
| On benefits | Yes or No. Unknown allowed | If unknown and found on a matched record, update |
| Country of origin | Must be a valid country code. File to be determined. Unknown allowed | If unknown and found on a matched record, update |
| Home language | Must be a valid home language. File to be determined. Unknown allowed | If unknown and found on a matched record, update |
| Disability | Yes, No or unknown. | If unknown and found on a matched record, update |
| Ethnicity | Must be a valid value (tbc). Unknown allowed | If unknown and found on a matched record, update |
| Pregnant mother | Yes or No. Unknown allowed | If unknown and found on a matched record, update |
| Expected due date | Either null or must be a valid date format, a valid date and a reasonable date (see specific date note below) | If missing and found on a matched record, update |
| Pregnant at registration | Yes or No. Unknown allowed | If unknown and found on a matched record, update |

### A.2 Child entity

|  |  |  |
| --- | --- | --- |
| **Field** | **Validation** | **Business rules** |
| Child ID | Autogenerated | Generated by the platform. ID design to be decided. |
| Primary carer email address | Must be a valid email address layout |  |
| Child DOB | Must be a valid date format, a valid date and a reasonable date (see specific date note below) |  |
| Sex of child | List value: Male, Female |  |
| NHS Number | Must be a valid NHS number, 10 digits and conform to Modulus 11 algorithm | If missing and found on a matched record, update |
| Original Postcode | Must be a valid postcode. | Always stored as indicator of a LEAP participant |
| Current postcode | Must be a valid postcode. | If different on a new matched record overwrite with new data |
| Original ward | Must be a valid London borough ward | Always stored as original residence |
| Current ward | Must be a valid London borough ward | If different on a new matched record overwrite with new data |
| Original borough | Must be a valid London borough | Always stored as original residence |
| Current borough | Must be a valid London borough | If different on a new matched record overwrite with new data |
| Date registered with LEAP | Must be a valid date format, a valid date and a reasonable date (see specific date note below) |  |
| Age at registration | Must be < 18 years |  |
| Country of origin | Must be a valid country code. File to be determined. Unknown allowed | If unknown and found on a matched record, update |
| Home language | Must be a valid home language. File to be determined. Unknown allowed | If unknown and found on a matched record, update |
| Disability | Yes, No or unknown. | If unknown and found on a matched record, update |
| Ethnicity | Must be a valid value (tbc). Unknown allowed | If unknown and found on a matched record, update |
| SEN | Must be a valid value. Unknown allowed. | If unknown and found on a matched record, update |
| Birth weight | Weight in Kgs | If unknown and found on a matched record, update |
| Full term | Yes or No. Unknown allowed | If unknown and found on a matched record, update |

## Appendix B - Standard values for validation

This section catalogues the set of standard values that will be used to validate the relevant data items.

|  |  |
| --- | --- |
| **Data item** | **Expected values** |
| Roles | |  | | --- | | Mother | | Father | | Other Primary Caregiver | | Other Family Member | | Child Minder | |
| Postcodes | Post code file for London |
| Wards | List of all London Wards |
| Boroughs | List of all London Boroughs |
| Employment status | |  | | --- | | Full time work | | Part time work | | In training or education | | Unemployed | | Full time parent | | Maternity Leave | | Other | |  | |
| Country | |  |  |  |  | | --- | --- | --- | --- | | 1 | United Kingdom | 19 | Bangladesh | | 2 | Nigeria | 20 | Germany | | 3 | Poland | 21 | Colombia | | 4 | Ghana | 22 | Albania | | 5 | Jamaica | 23 | Eritrea | | 6 | Somalia | 24 | Ecuador | | 7 | Portugal | 25 | Côte d'Ivoire | | 8 | Italy | 26 | Lithuania | | 9 | France | 27 | Viet Nam | | 10 | Brazil | 28 | Russian Federation | | 11 | Sierra Leone | 29 | South Africa | | 12 | United States of America | 30 | Algeria | | 13 | China | 31 | Ethiopia | | 14 | Romania | 32 | Philippines | | 15 | Australia | 33 | Turkey | | 16 | India | 34 | Bulgaria | | 17 | Ireland | 35 | Other | | 18 | Spain |  |  | |
| Language | |  |  |  |  | | --- | --- | --- | --- | | 1 | English | 16 | Albanian/Shqip | | 2 | Portuguese | 17 | Benga | | 3 | Somali | 18 | Lingala | | 4 | Spanish | 19 | Tigrinya | | 5 | French | 20 | Igbo | | 6 | Yoruba | 21 | German | | 7 | Akan/Twi-Fante | 22 | Urdu | | 8 | Amharic | 23 | Turkish | | 9 | Arabic | 24 | Vietnamese | | 10 | Polish | 25 | Russian | | 11 | Chinese | 26 | Bulgarian | | 12 | Amarakaeri | 27 | Swahili/Kiswahili | | 13 | Bengali | 28 | Hindi | | 14 | Italian | 29 | Kurdish | | 15 | TGR | 30 | Other | |
| Ethnicity | |  |  | | --- | --- | | Unknown | Mixed Other | | African (excluding Somali) | Mixed White/Asian | | Somali | Mixed/Black African | | Any Other Group | Mixed/Black Caribbean | | Asian Other | Pakistani | | Bangladeshi | Portuguese | | Black Other | Turkish | | Caribbean | Vietnamese | | Chinese | White British | | Greek | White Irish | | Gypsy/Roma | White Other | | Indian |  | |
| SEN | |  | | --- | | Autistic Spectrum Disorder | | Behaviour, Emotional & Social Difficulties | | Hearing Impairment | | Moderate Learning Difficulty | | Multi-Sensory Impairment | | Other Difficulty/Disability | | Physical Disability | | Profound & Multiple Learning Difficulty | | Severe Learning Difficulty | | Specific Learning Difficulty | | Speech, Language and Communication Needs | | Visual Impairment | | Not applicable | |