South London and Maudsley NHS
NHS Foundation Trust

# Biomedical Research Centre

## eLIXIR Project

OPERATIONAL SECURITY MODEL V1.2

Development and change log:

| Date | Version | Presented to | Outcome/changes made |
|---|---|---|---|
| 22.03.2018 | 1 | Research Ethics Committee and HRA | Approved |
| 15.02.2019 | 1 | Oversight Committee | Approved |
| 31.09.2019 | 1.2 | Oversight Committee | Approved |
| | | | |
| | | | |

## 1. Introduction

1.1 The eLIXIR partnership, funded by the MRC in 2017 is a multidisciplinary academic collaboration which aims to combine maternal, infant and child health data into a single resource to allow information from large numbers of mothers, babies and children to be studied, ultimately over a long period of time, in order to better understand the life course influences on mental and physical health.

1.2 The eLIXIR platform seeks to provide a mechanism for research datasets to be assembled from linked data from clinical records across King's Health Partners sites, including linkages between mothers and offspring, under appropriate and approved (by Caldicott Guardian) levels of anonymity and data security. It will also be designed to allow potential future incorporation of other relevant data, health and otherwise. This document details the technical and procedural elements in place to safeguard the legal and ethical rights of service users during the development and use of the eLIXIR platform.

1.3 Technical elements of the model draw strongly on experience gained in setting up the Clinical Record Interactive Search (CRIS) data resource at the Maudsley NIHR Biomedical Research Centre, which has received extensive consideration and full approval from the Trust Executive, CRIS Oversight Committee and independent Research Ethics Committee review (Oxford REC C 08/H0606/71+5).

## 2. eLIXIR – Background and Rationale

2.1 Long term population and cohort studies have proven invaluable in understanding the biological mechanisms of adult diseases. The suggestion that 'long term' should now include data from the earliest stages of life follows more than two decades of research proposing that early life 'exposures' in utero and in early postnatal life are independently associated with risk of physical or mental health disorders in childhood and in later life. Pharmacologically induced teratogenicity is well recognized. Less obvious environmental insults to the embryo and early fetus during stages of developmental plasticity leading to persistent effects include the biological consequences of maternal nutritional excess or deficit, stress and mental health disorders, endogenous or exogenous glucocorticoids, Caesarean section, and pre-eclampsia. Environmental pollutants may also have insidious long-term influences. Not only are there lifelong consequences for the child; common complications in pregnancy including pre-eclampsia and gestational diabetes are associated with longer term maternal morbidity. For the neonate the biological stresses of premature delivery, inappropriate nutritional strategies, infection, and exogenous steroids, have also been implicated in sub-optimal childhood health and development, a prelude to adult physical and mental health disorders. During childhood the potential for adverse influences of the environment, especially health and social factors on the immediate and longer-term health of the child is also well established.

2.2 In the UK, health data exists in a lot of different record systems, and information from mothers and babies are not routinely linked apart from in records at or around the index childbirth. The eLIXIR project aims to provide a linkage resource to bring together clinical data sources across King's Health Partners (KHP) (an Academic Health Science Centre comprising three NHS Trusts (Guy's and St Thomas', King's College

Hospital, South London and Maudsley) and a university partner (King's College London)). The clinical data include maternity records, fetal scan records, neonatal intensive care records, and children's hospital data. This will support 'observational research' into many disorders and other factors measurable across the life course and trans-generationally and provide a step change in UK research capability in the life course of physical and mental health disorders. eLIXIR is recognised to be a large and potentially complex project.

2.3 The eLIXIR platform will initially support the linkage of clinical datasets derived from and owned by KHP services. These data sources include the Clinical Record Interactive Search (CRIS) at the South London and Maudsley NHS Foundation Trust (SLaM) and sources of maternity, neonatal and paediatric data from Guy's and St Thomas' (GSTT) and King's College Hospital (KCH) NHS Foundation Trusts. BadgerNet Maternity and BadgerNet Neonatal systems will be the initial sources of maternity and neonatal data. These are separate data banks that are not linked with each other for research purposes, although clinical data are shared. Use of clinical datasets and record linkage between BadgerNet and CRIS will provide rich clinical data and an early stage example of linkage between mothers and babies that address some of the significant methodological issues within current data resources. For example, recruited samples in clinical studies are frequently (if not inevitably) unrepresentative of their source population, and large administrative datasets (such as those from Scandinavia) currently lack detail, for example on information such as medication exposures in pregnancy. As well as making efficient use of existing resources the proposed study will allow more representative data from the local population that KHP serves, particularly from disadvantaged and vulnerable individuals whose experiences are often the most challenging to obtain information on, as well as those who may be too unwell to participate in clinical studies. These data will provide information about what is happening in health services in a real-world setting and will enable long term follow-up minimising loss to follow-up and the expense of large population cohort studies.

## 3. eLIXIR Security

1.1 To fulfil its potential to improve outcomes and opportunities for services users it is essential that the legal and ethical rights of service users are safeguarded. With this in mind the development of the eLIXIR security model is a core element of the overall eLIXIR development project. eLIXIR security is broken down into two component parts – technical specifications that were built into the platform itself during development; and procedural standards that govern the launch and day-to-day use of eLIXIR.

1.2 eLIXIR Security – Technical Elements: these have developed to mirror, as closely as possible, those developed for CRIS by the BRC Working Group alongside functional requirements during the specification phase of the development project, from September to November 2007. The following table lists headline security elements included in the CRIS design. For detailed specifications see Appendix A.

| Table 3.2.1 Headline Elements of eLIXIR Technical Security | |
|---|---|
| **i.** | **SEARCHING AGAINST PERSONAL IDENTIFIABLE INFORMATION (PII)**<br>All PII should be removed from eLIXIR platform data repositories entirely, including references in text and dedicated PII fields, or sufficiently truncated/modified to protect confidentiality, e.g.:<br>• Date of birth: truncated to month and year of birth only<br>• Postcode: modified to Lower Super Output Area<br>• Ethnic category: collapsed into the NHS standard 16+1 categories. |
| **ii.** | **PSEUDONYMISING SERVICE USER IDS**<br>A unique ID number is created for all service user tables in eLIXIR. All linkage tables are separated from the searchable eLIXIR data repositories, and so are unavailable to research users. |
| **iii.** | **DEANONYMISING**<br>It will not be technically possible for eLIXIR research users to reveal the link between project IDs and any source system ID / NHS number or vice versa under any circumstances. Only the eLIXIR System Administrator will be able to access these links. |
| **iv.** | **ACCESS CONTROL**<br>Access control to eLIXIR repositories will be password protected. |
| **v.** | **AUDIT TRAIL**<br>All eLIXIR activity will be logged in an audit trail accessible to the eLIXIR Administrator only. |
| **vi.** | **OPT-OUT**<br>Patients have the right to opt out from having their record in the eLIXIR platform. The eLIXIR platform includes an 'opt-out' function that automatically excludes records specified in a configurable exclude list. |

1.3 eLIXIR Security – Procedural Elements: these again have been developed to mirror as closely as possible those developed by the CRIS Security and Confidentiality Procedures Development Group, a time-limited project team chaired by Dr Felicity Callard, BRC Stakeholder Participation Theme. Membership included the Trust's Caldicott Guardian and child protection lead. The group agreed that the following must be in place or have taken place before CRIS could be formally launched, and the same will be true for eLIXIR.

| Table 3.3.1 Processes and actions required for eLIXIR to be fully implemented | |
|---|---|
| **i.** | The eLIXIR Oversight Committee is responsible for overseeing and monitoring the use of the eLIXIR platform, including:<br>Managing the eLIXIR application process. All projects proposing to use eLIXIR are required to submit a written application to the committee. Applications are judged according to:<br>• underlying value and potential benefits of the project, e.g. to inform patient care;<br>• appropriate supervision/ governance, e.g. research governance for research projects; formal clinical governance approval for audits; Trust director sign-off for service evaluation;<br>• inadvertent risk of deanonymisation, e.g. the likelihood of particularly small cohort/ cell sizes (< 10 cases); appearance of high profile publicly known/ published information, etc. In these cases, additional measures may be put in place to safeguard confidentiality.<br><br>Note: The eLIXIR platform has not been developed to support service management. Applications to use eLIXIR to evaluate or monitor staff-level performance will not be granted. Applications to use staff names for legitimate research / audit purposes may be granted but additional supervision may be put in place.<br><br>**Monitoring use -** e.g. comparing intended and actual use of the eLIXIR platform through routine monitoring of the audit trail;<br>**Managing the eLIXIR Security Model** – ensuring the model is fully implemented at all times and updated as required to meet appropriate standards.<br>**Managing the eLIXIR Communications Plan** – ensuring relevant stake-holders, including service users and staff, are able to access relevant information about eLIXIR, including the service user's right to opt-out<br>**The eLIXIR Administrator** - acts on behalf of the Oversight Committee, including managing eLIXIR applications, users' accounts and access to audit logs, committee meetings.<br><br>The eLIXIR Oversight Committee is accountable to KHP's Caldicott Committee and includes Caldicott representation. |
| **ii.** | Individual users named in approved projects require a KHP Trust contract (honorary or substantive) or research passport, which ensures users are contractually obliged to adhere to relevant Trust policies regarding confidentiality and data protection. Anyone wishing to use the data who is not an employee with a KHP Trust must obtain a substantive or honorary contract with a KHP trust. |
| **iii.** | Patient identities from Trust Audit and Service Evaluation projects carried out on the eLIXIR platform may not be revealed without explicit Trust Caldicott approval, in addition to eLIXIR Oversight approval. In these cases, reverse searches are carried out in batch by the eLIXIR Administrator. The link between the ID and the patient identifier is not revealed to dataset users. |
| **iv.** | The eLIXIR platform may be used to extract clinical data for existing research cohorts recruited from the source Trusts. In these cases, the IDs of participants will need to be linked to known identifiers, i.e. source system ID or first name/last name/DOB combination. Permission to link known identifiers to eLIXIR platform data will only be given if explicit consent to access records has already been given as part of an existing, ethically approved project. In these cases, reverse searches are carried out in batch by the eLIXIR Administrator. The link between any eLIXIR ID and the patient identifier is not revealed to dataset users. |
| **v.** | By default, patient level data must remain at all time within the SLaM firewall, including the data build phase; searchable eLIXIR data repositories; any table linking eLIXIR IDs with source-system and other identifiers, front-end applications to access these data, and all patient-level |

| | |
|---|---|
| | data exported from the eLIXIR. This ensures these data are subject to the same rigorous security standards (technical and policy) applied to other patient level data by the Trust.<br>In addition, eLIXIR data should not be loaded onto Trust data sticks including encrypted sticks. All eLIXIR users will be made aware of this decision when access is first granted. |
| **vii.** | The technical security elements are fully functional including the deidentification algorithms. |
| **viii.** | The eLIXIR platform has appropriate, up-to-date formal approvals, including Trust (Caldicott and Trust Executive) and ethics approval as an anonymised data source for secondary analysis. |

**Note: Elements of this model may be overridden by the eLIXIR Oversight Committee:**

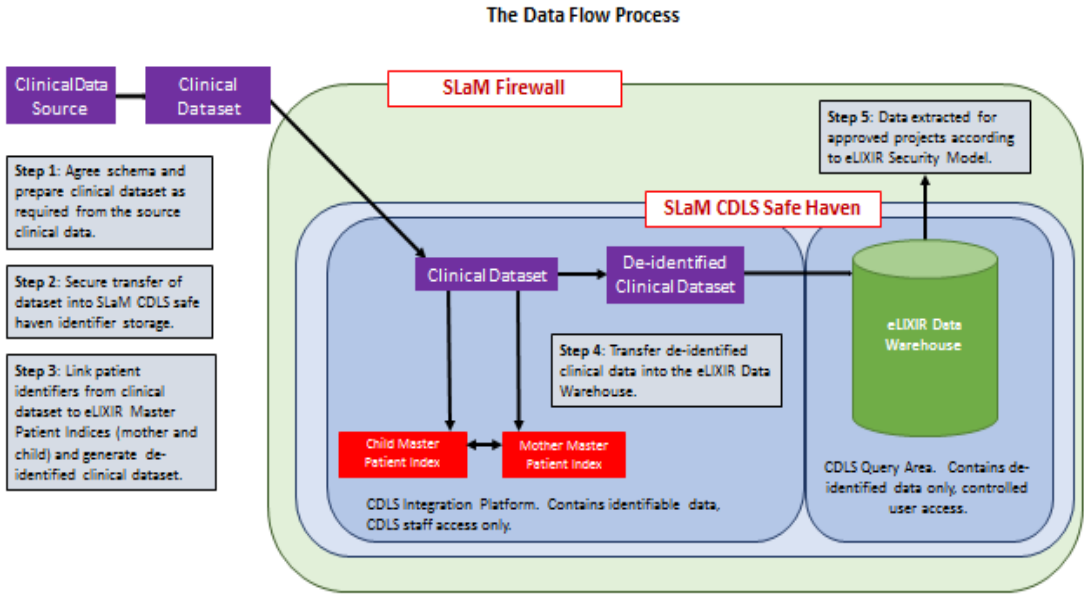
i. In carrying out their day-to-day duties the eLIXIR technical team is exempt from eLIXIR access rules. This team includes SLaM Clinical Data Linkage Service (CDLS) IT technical support, eLIXIR administration and contracted third-party support (subject to stringent approval standards). In all cases there will be a contractual responsibility to protect identifiable patient data.

ii. Specific projects may request patient level data to be taken out of the firewall, e.g. linking eLIXIR data with other health data outside SLaM through a trusted third party. In these cases, applications detailing project-specific procedures to protect data must be made and require prior and explicit approval from the KHP Caldicott and eLIXIR Oversight Committees. Subject to SLaM Caldicott recommendation, Section 251 approval may also be required.

## 4. eLIXIR Data Processing Model

Suppliers of the data, e.g. SLaM, GSTT, and KCH, are Data Controllers in common with regards to the linked eLIXIR data.

In addition to SLaM being the data controller of the CRIS data, SLaM is acting as the data processor for the eLIXIR platform. The SLaM Clinical Data Linkage Service (CDLS) manages the platform, including receiving and linking source data using Patient Identifiable Information (PII), secure hosting and managed access to the de-identified eLIXIR data.



# Appendix A: eLIXIR Technical Security Model

## 1. HOSTING

The eLIXIR repositories will only be available from inside the SLaM firewall. The eLIXIR platform will be hosted within the SLaM CDLS data centre and administered by SLaM CDLS IT support service. eLIXIR installations are protected by relevant SLaM IG and IT Security Policy and practice. SLaM has formal IG Toolkit approval.

## 2. DATA DE-IDENTIFICATION

Data in eLIXIR repositories are fully deidentified.

### 2.1 Pseudonymisation

eLIXIR IDs are pseudonyms created for every record as it is passed into an eLIXIR source table. These IDs are locally generated against a combination of identifiers, including source system ID, NHS number, first name, last name and date of birth. The IDs are not fully anonymised (i.e. the link between the IDs and source Trust IDs is not permanently destroyed), so that the ID allocated to each record remains consistent over time whilst allowing routine updates from the source and periodic full data rebuilds.

However, the link between the eLIXIR-generated IDs and patient identifiers is not retained within searchable eLIXIR repositories and so is not accessible by eLIXIR platform users, rendering eLIXIR IDs as effective anonyms for eLIXIR users.

### 2.2 Removal of Personal Identifiers

All personal identifiers (PIs) are removed from eLIXIR repositories to protect the identity of services users.

**Structured and small-text data:** dedicated fields from source Trust data where PIs are recorded, e.g. first name, last name, address details, date of birth etc. are either excluded entirely from the searchable eLIXIR repositories or are truncated within eLIXIR.

**Truncation/ modifications:** particular PIs are truncated so they are available for research purposes, without compromising confidentiality, e.g.

- Date of birth – from dd/mm/yyyy to 01/mm/yyyy
- Postcode – Lower Super Output Area is derived from full postcode and replaces postcode in all repositories. Output Area is also derived from full postcode but is only available to the eLIXIR administrator to extract deprivation scores and other area-based census data on behalf of eLIXIR platform users.
- Ethnic category collapsed to NHS standard 16+1

N.B. Date of death is not truncated or modified in eLIXIR.

**Unstructured open text**: service-user forename, surname, full date of birth (any known format), full address (any of lines 1-3 + postcode), phone numbers, alias(es) are all masked in unstructured field returns, e.g. Mr John Smith will be shown as Mr ZZZZZ ZZZZZ. Equivalent information relating to contact(s) – forename, surname, full address (any of lines 1-3 + postcode) and phone numbers will be masked by 'QQQQQ'. Where PIs are shared by patient and contact, e.g. last name, ZZZZZ will be used.

Detailed evaluation has demonstrated the effectiveness of the CRIS masking algorithms (Fernandes et al. [2013] Development and evaluation of a de-identification procedure for a case register sourced from mental health electronic records, BMC Medical Informatics and Decision Making.2013, 13:71).

This functionality is currently only available for SLaM data in the CRIS repository, but the functionality will be a condition of incorporation of any text from other sources in the eLIXIR platform.

## 3. THE eLIXIR PLATFORM

The following technical security components will be built in to the application.

### 3.1 Role-Based Access Control

Two different roles are available:

**Role 1:** Observer – can:

- Observers can carry out full search and export from deidentified eLIXIR index
- Observers cannot conduct reverse searches (access the link between the pseudonyms and related patient identifiers) or access email for recruitment

**Role 2:** System Administrator - can:

- create and modify user profiles, including allocating authorisation to use the eLIXIR platform and modification of user details
- view and truncate the Audit Trail
- manage the list of ethically approved projects in eLIXIR
- Full search

### 3.2 Audit Trail

The eLIXIR platform will log details of all searches carried out, including:

- login, logout date/ timestamp
- search instance details
- user name, date timestamp
- chosen search parameters and search parameter values
- chosen result dataset outcome variables

### 3.3 Access Controls

Access through the eLIXIR platform is password protected using authenticated SLaM network user name and password. It is not possible for data exported from the eLIXIR platform to be saved directly outside the firewall.

### 3.4 Sign-off

All functionality in the eLIXIR platform relating to security will be tested and signed off by a Project Security Group following its development.

## 4. OPT-OUT

During the eLIXIR build and for all data updates, records will automatically be excluded from the eLIXIR repositories if they appear on an exclusion list of source system IDs.

## 5. DATA RETENTION

eLIXIR data files used for research projects will be kept for 10 years after the final publication in accordance with standard research governance. Once eLIXIR projects are completed the datasets will be archived and then destroyed 10 years from the date of archiving. All data will be disposed of in accordance with the SLaM Confidential Waste Procedure and the Information Security Policy. Physical servers will be destroyed by a SLaM Trust approved WEEE (Waste Electrical and Electronic Equipment) Disposal Company who will provide SLaM with a certificate of destruction. The raw databases held by the CDLS will remain anonymised and within the SLaM firewall at all times.