**DISCOVERY PROGRAMME**
**Annual Report – January 2019**

## 1. BACKGROUND

In January 2016, a Project Initiation Document (PID) for the Discovery Programme was prepared, setting the objectives, deliverables, benefits and initial stage governance.

The Discovery Programme is currently being implemented to establish, deliver and ultimately manage a secure data service with linked combined identifiable data from all systems supporting direct health care in east London. In its initial phase, the programme has focussed on primary and acute care for the Waltham Forest, East London & City (WELC) CCGs and the Transforming Services Together (TST) Programme, but the aim is to progress to include community health services, mental health and social care and to extend the geographic reach of the dataset to support the work of the full north east London sustainability and transformation plan – East London Health & Care Partnership (ELHCP).



**Key**

- Boroughs covered by *Transforming Services Together* (TST)
- Remaining boroughs covered by the Sustainability and Transformation Plan (STP)

**Boroughs in the NE London Footprint**
1. City of London
2. Tower Hamlets
3. Newham
4. Waltham Forest
5. Hackney
6. Redbridge
7. Barking & Dagenham
8. Havering

**Hospitals with A&Es in the TST boroughs**
H1. The Royal London Hospital
H2. Newham University Hospital
H3. Whipps Cross University Hospital
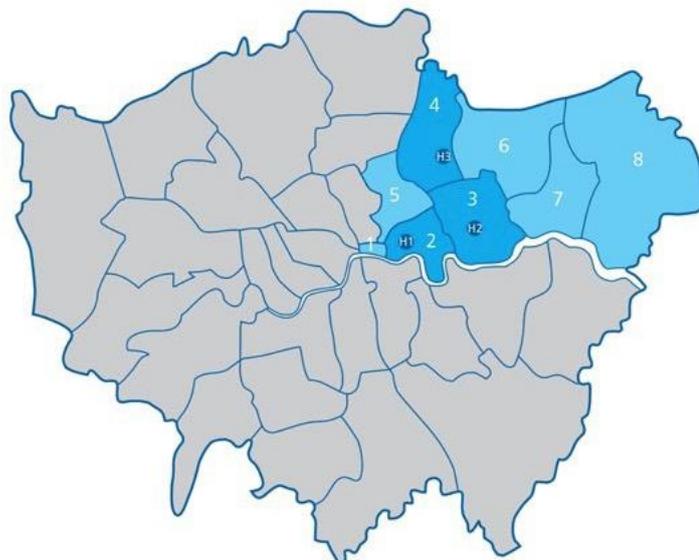
This programme complements and enhances existing systems. It does not replace them and has four main aims:

1. To predict, anticipate or inform **individual** health needs from algorithms running in real time (or as near as possible) and to deliver the insight gained directly into the patient's record across the whole of their pathway, whether in primary or secondary care or elsewhere, thus creating the opportunity to improve or prevent adverse outcomes.

2. To expand the existing primary care informatics driven **population health** programme in east London, led by the Clinical Effectiveness Group (CEG) at Queen Mary University of London, to all health and care sectors.

3. To enable the real time **reporting on programmes** by providers and commissioners supporting clinical improvement and new payment mechanisms. This would involve reporting on a depersonalised or identifiable cut of the clinical data, as appropriate.
4. To use data by third parties (commissioners, public health, and academics) **to support research, development and planning**, whether on consented identifiable data, or the depersonalised dataset. East London would thus become a research-enabled community.

The Discovery Programme governance architecture was originally shaped as below:
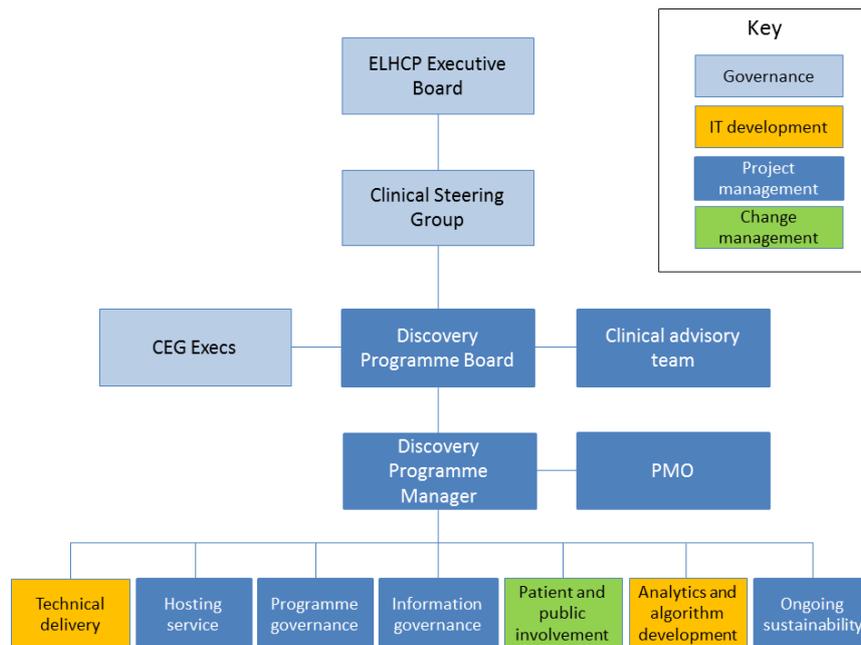


This Annual Report is structured in line with these original workstreams and seeks to provide a status update against each of the seven workstreams identified, as well as more detail on the progress of priority work elements.

## 2. TECHNICAL DELIVERY

The Discovery Programme has achieved technical accreditation to access data from Primary Care via the GP Systems of Choice (GPSoC) Assurance Steering Group program. Funding is available to accredited subsidiaries until the end of the GPSoC-R contract.

Data is now being received, via the GPSOC contract for Practices using EMIS and Vision. We have passed the Pairing Process stage for engagement with TPP and Microtest, however those suppliers are yet to make the service available. In the interim we are working with local organisations to access TPP data.

There follows an overarching description of the data feeds completed or in development, categorized into the following care sectors: Primary Care; Secondary Care; Integrated Urgent Care; Community; Extended GP Care.

# Primary Care

All the East London practices (170) in City & Hackney, Tower Hamlets, Newham, and Waltham Forest are now publishing their data into the Discovery Data Service (DDS).  The two practices, in Waltham Forest CCG, using the TPP system are expected to be live in the DDS within the next month.

|  | Total Practices | Activated as at Oct 2017 | Activated now, at Jan 2019 |
|---|---|---|---|
| City & Hackney | 43 | 34 | 43 |
| Newham | 53 | 35 | 53 |
| Tower Hamlets | 35 | 35 | 35 |
| Waltham Forest | 39 | 15 | 39 |
| TOTAL | 170 | 119 | 170 |

# Secondary Care

*Barts Health NHS Trust*

**Cerner Data - Received:**
ADT messages via HL7 (real-time)
Power Insight Data Warehouse extract, including clinical events and documents (daily)
Inpatient, Outpatient, A&E SUS files (daily)
Procedure, Problem and Diagnosis CDS files (daily)
Pregnancy, Birth and Delivery CDS files (daily)
HL7 Data – requires further development work, expected to be complete by end March-19.

*Homerton University Hospital NHS Foundation Trust*

**Cerner Data - Received:**
ADT messages via HL7 (real-time)

# Integrated Urgent Care

*Adastra*

1 organisation (2 Urgent Care centres) have published their data to the DDS.
Orders have now been placed with Advanced (Adastra) for 5 East London organisations, each containing 2 or 3 services, and the Hurley Group Extended Services.

## Community

*Barts Health NHS Trust*

>    *EMIS  -  All Barts Community Services are live in the DDS*

*East London Foundation Trust*

>    *EMIS  –  3 services live (data streams merged to report as a single entity's activity)*
>    *RIO    – Continuing to engage with suppliers and providers to bring this data in.*

## Extended GP Care

>    Two users of EMIS Web for Federated and Extended Care are live in the DDS.

### 3.    HOSTING SERVICE

The Discovery Programme is committed to protecting patient data, whilst making it available to meet the objectives of the programme. Over the past few years, there have been considerable advances in data hosting services. The programme has now updated the data hosting environment to maximise on this opportunity to deliver a more efficient and cost effective service.

The Discovery Data Service is hosted in a secure cloud environment which is approved by NHS Digital. High levels of security, availability, and scalability, combined with a Health and Social Care Network-approved connection, provides an exceptional service.

Data published into Discovery is controlled by strict data sharing processes and agreements. Data is currently stored in a fixed hosted environment which is an N3 aggregated solution with IG level 3 accreditation to IGSOC and ISO27001 standard and which uses 2-factor authentications.  Cloud services are made available to subscribers on approval from the Discovery Programme Board. Data remains at rest on servers hosted by the data publishers, with authorised subscribers accessing data information as clients.

### 4.    PROGRAMME GOVERNANCE

As part of its delivery, the Discovery Programme has established formal governance mechanisms to ensure that the service is managed by all contributing members to the benefit of all contributing members. A joint Board, with clinical and management representatives from all organisations that contribute data, governs both the programme and organisations' access to data. The Discovery Programme Board's holding organisation is NHS Tower Hamlets CCG (as host for East London HCP), and the programme office operates from CEG.
During 2018 we have built relationships and formalised the sustainable working model for the programme, initially focussing on bringing together representatives from the central contributors to

the Discovery Programme Board. This step aims to ensure that we remain inclusive enough to be fit for the future involvement of broader stakeholders (for example local authorities) whilst clearly articulating the specific benefits to be realised by member organisations.

The Discovery Programme is underpinned by the Discovery Working Group which meets fortnightly and co-ordinates the day-to-day activities of Discovery Programme staff. These meetings focus on project and technical co-ordination, and also act as a gateway mechanism to consider new and progressing applications for intelligence and insights from the data, and requests from third parties for data access.

## 5. INFORMATION GOVERNANCE

A framework for reviewing applications to access and use data curated within the Discovery Data Service was approved by the Discovery Programme Board in June 2018.  The aim is to establish a transparent process for reviewing and prioritising those applications judged to provide most benefits to individuals and/or the communities we serve, and thereby to provide evidence to them of the trustworthiness of the Discovery Programme in the uses of Discovery data and to minimise any reputational risks to the Programme.  The legal basis for allowing access to and use of data by third parties not for the direct clinical benefit of a consented patient was reviewed in 2016 by Dame Caldicott, National Data Guardian for Health and Care, and is subject to the Data Protection Act 2018.  The process set out for the Discovery Programme adheres to established legal and ethical considerations and requirements.
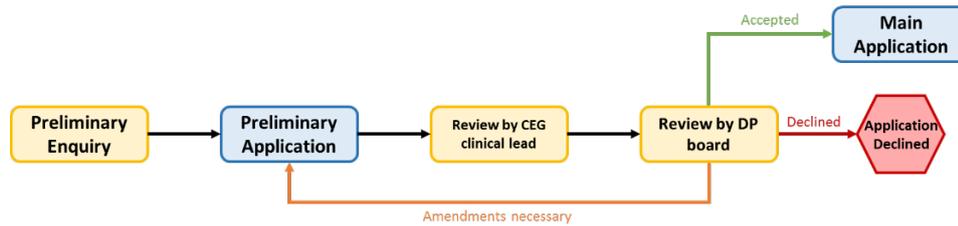
As part of this framework, a Discovery Programme Public Interest, Data Privacy and Access Review (PIPA) Panel is being established which will be responsible for advising the Discovery Programme Board on applications to use Discovery Programme data in line with these aims and purposes, including medical device appraisals where these include use of patient data.  The process is designed to consider all applications, whether from the NHS (including for commissioning and business intelligence), academics and researchers, or from commercial or third sector users. These uses form part of the benefits realisation anticipated in establishing the Discovery Programme in east London. The PIPA Panel will support the Discovery Programme by reviewing and advising on applications to use Discovery Programme patient data, and considering that these uses should
- Safeguard the privacy, anonymity and confidentiality of individual patient data
- Maintain the trustworthiness of the Discovery Programme, the NHS and other organisations providing their data to the Discovery Programme
- Maximise public benefit and the fair distribution of these benefits to patients, the public and the NHS
- Add value to the NHS and its provider organisations through use of patient data assets
- Have relevant scientific and ethics approval where appropriate
- Meet the principles guiding access as set out in the Discovery Programme Public Interest, Data Privacy and Sharing Policy
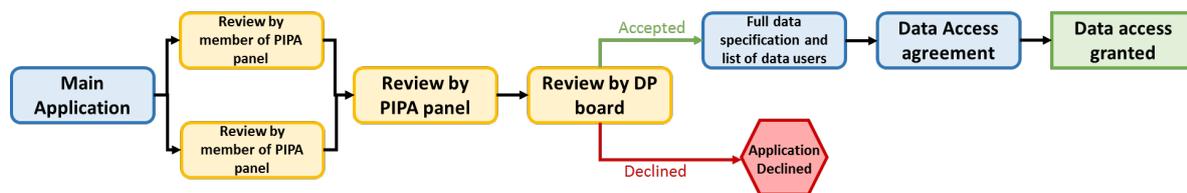
 The Panel will report its recommendations to the Discovery Programme Board which is responsible for all decisions on data access. The preliminary and main application process (outlined in Figure 1) is being implemented in an electronic data capture system (RedCap) to enable an audit trail on all applications and decisions: this will be piloted and go live in the first half of 2019.

Figure 1: Discovery new enquiry/application process flow

a) Preliminary Application

Preliminary Enquiry → Preliminary Application → Review by CEG clinical lead → Review by DP board

Review by DP board — Accepted → Main Application

Review by DP board — Declined → Application Declined

Amendments necessary (Review by DP board → Preliminary Application)

b) Main Application

Main Application → Review by member of PIPA panel / Review by member of PIPA panel → Review by PIPA panel → Review by DP board

Review by DP board — Accepted → Full data specification and list of data users → Data Access agreement → Data access granted

Review by DP board — Declined → Application Declined

Data are to be hosted on data platforms that conform to the Discovery Programme's expectations in relation to the legal and ethical basis for data sharing and in line with current legal and ethical provisions. Projects using unconsented pseudonymised data will be hosted within an approved Discovery data environment, with access permissions created for approved and accredited Users. Projects based on consented research studies can be hosted in other data environments provided these comply demonstrably with specified standards of cybersecurity and information governance as well as the stipulations of the Data Access Agreement.  We are currently exploring options for suitable data platforms.

6.      PATIENT & PUBLIC INVOLVEMENT

The work to appropriately involve and engage with the NEL population has now been defined and is being procured. By partnering with the Wellcome Trust-hosted programme Understanding Patient Data (UPD), and with the NEL STP, the Discovery Programme will be initiating further conversations with Londoners about data and privacy considerations that are relevant to how Discovery data is used.

In summary, this will involve the following:

- Preparation of appropriate engagement materials to articulate the scope and purpose of the conversation
- Borough-based public engagement events and staff engagement activities
- Local insight collection, with particular attention to identifying and engaging with seldom-heard communities across the NEL geography
- An open and productive conversation with equity between members of the public and professions, professionally-led and communicated
- Collation of viewpoints as outputs, to provide learning insights back to leadership and stakeholders of the Discovery Programme

The approach is aligned with existing work currently progressing or planned as part of the One London Local Health and Care Record Exemplar (LHCRE) and Health Data Research UK (HDRUK)[1] with whom we work closely. However, the work procured would be specific to the Discovery Programme, recognising the particular needs for clarity when communicating about how Discovery data and information is used.

The proposed timeline is to sign a contract early in 2019, deliver conversation planning details in April 2019, and to hold events with groups of Londoners throughout the summer. The resulting material outputs would be available at the project's close in July or August 2019. The material outputs would include recommended language choices, tailored presentational materials, public information notices, each designed to maximise patient and public understanding and acceptance. Materials would be available from August 2019 onwards and provide a solid framework from which to perform future engagement events, across London and in other locations.


7.      **ANALYTICS AND ALGORITHM DEVELOPMENT**

As the Discovery Programme has developed there have been increasing requests for Discovery data and information utilities.

There are a number of steps towards achieving data access and information utilities and algorithms. These start with initial conversations between the Discovery team and stakeholders, and progress by following the approval process (as described in section 6 above) including seeking approval by the Discovery Programme Board. An approval authorises the specific use of extracted data and analysis for the purposes stated in the application, which may include the development and use of new utilities and algorithms.

---

[1] https://www.hdruk.ac.uk/

## *Integral Discovery Utilities (Live)*

**111 Frailty Flag and London Ambulance Services**

The Discovery data service **frailty flag** is now live in an adoption and roll-out phase. Figure 2 shows the process flow for a call to 111 involving the flag.

This is the first time that Discovery has been able to place new knowledge (Frailty flag) derived from diverse clinical systems into another clinical system (NHS 111/LAS Adastra system). The frailty flag alerts the call handler to triage the call for more rapid clinical review. There are 6148 severely frail patients flagged within 1.4 million individuals in the current Discovery dataset.
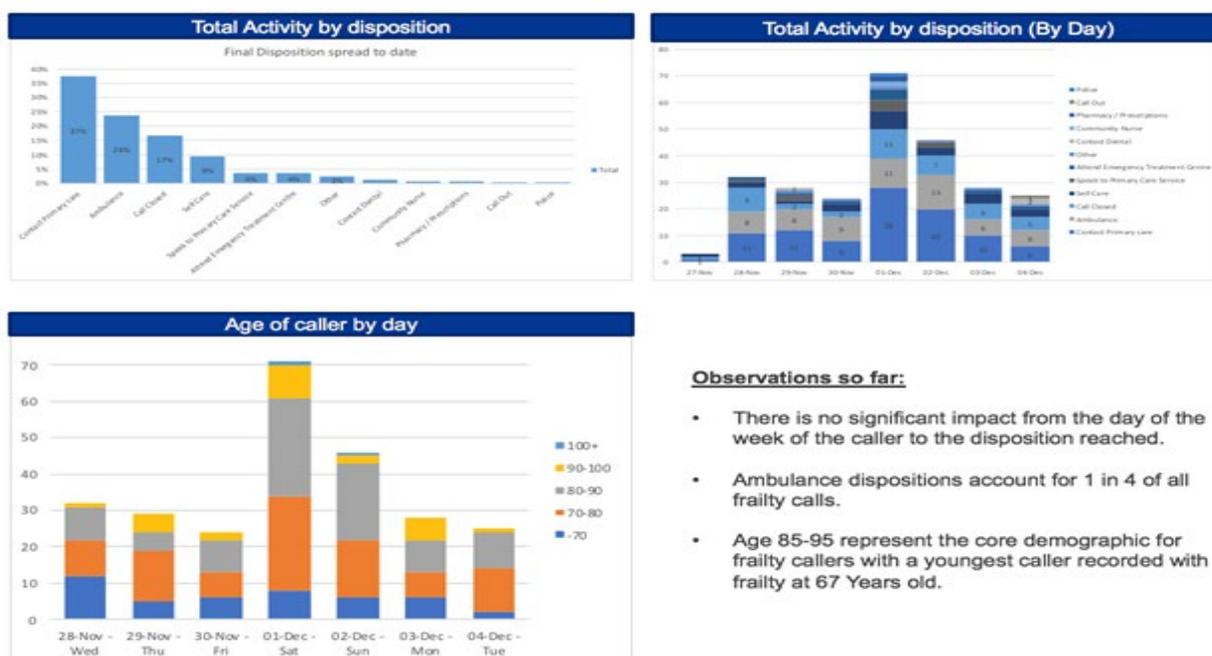
Figure 2: Process flow for a call to 111 involving the Discovery Frailty Flag
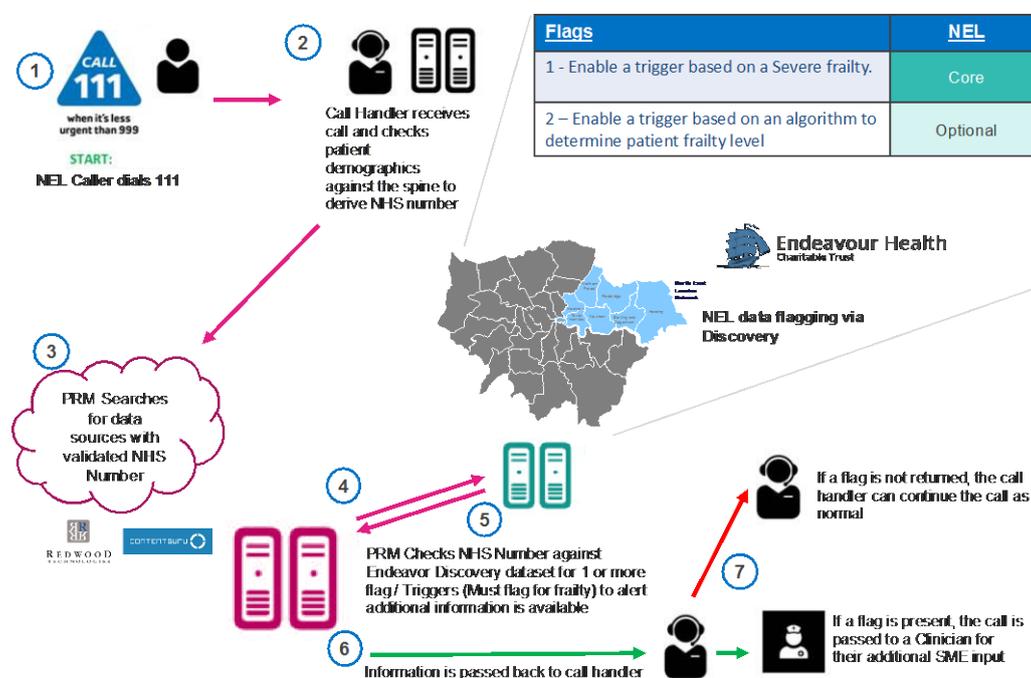


This service is potentially scalable across London as part of the One London LHCRE initiative and nationally.

Initial analysis of the frailty calls has started (Figure 3) and development and analysis is ongoing. Figure 3 shows an internal dashboard being used by the Discovery team to monitor and analyse calls to 111 which make use of Discovery Frailty flag utility. Total Activity is being monitored by disposition, across days, and split by age of caller. These charts demonstrate the types of information and insight which can be derived by integrating patient demographics and clinical data drawn from Discovery when applied as a utility within the 111 service.

Figure 3: Preliminary Frailty calls analysis



Observations so far:

- There is no significant impact from the day of the week of the caller to the disposition reached.

- Ambulance dispositions account for 1 in 4 of all frailty calls.

- Age 85-95 represent the core demographic for frailty callers with a youngest caller recorded with frailty at 67 Years old.

## *Integral Discovery Utilities (In Development)*

**Clinical Effectiveness Group (CEG) Discovery data assurance**

Users within CEG will increasingly access the Discovery database for the purposes of on-going data checking and validation between the existing CEG database and Discovery. Especially when new data is published into Discovery, assurance will be a key step.

**Where is my patient? Who is providing care?**

A utility to identify, for any patient, their current geographical location and the responsible and involved care professionals. The tool would give complete information on both the organisational setting (i.e. is the individual in hospital, at a GP Practice, at home under community care, within an out of hours setting etc.) and also of the health and care professionals delivering care. Such a utility would enable better integration and communication links between the patient and all of the health and social care professionals from whom the patient receives care.

**Population health manager**

> This is under development to clarify the 'concepts' required to define a condition or set of conditions: for example, the concept 'asthma' might comprise half a dozen different codes using a number of different code sets (READ/Snomed/ICD10/OPCS).

**Geographical identifiers**

> We are developing an address linkage algorithm and service to assign Unique Property Reference Numbers (UPRNs) from the industry standard Ordnance Survey AddressBase Premium (ABP) property gazetteer to NHS patient addresses held in Discovery. ABP is derived from local authority-maintained address management systems, the Royal Mail Postcode Address File, and Ordnance Survey data, is quality-assured by GeoPlace, and available to the NHS under the Public Sector Mapping Agreement at no extra cost.  This will give unprecedented geographical 'resolution' for de-identified datasets (currently limited to groupings of around 1500 individuals in Lower Super Output Areas). This more granular resolution is a step-change in NHS data capability. This development is a collaboration between Discovery and Dr Gill Harper, a HDRUK-funded post-doctoral research fellow (CEG, QMUL).

## *Research Studies*

Several projects, given Board approval, are at varying stages of development and these are summarised below.  In addition, several applications currently under discussion are outlined in the Project Register**.** We will seek to progress those that are approved by the Board when access to the Discovery data becomes available.  These are likely to increase in number as Discovery goes live and, given the Discovery team time and resources required to review and deliver on each application, a process will be required for the Programme to prioritise those projects that provide most benefits to the communities we serve.

Summary of studies underway

**REAL-Health** *(PI: Robson)*

> The study is funded by Barts Charity and started in October 2018. It includes three themes:
> - Child Health (Dezateux)
> - Respiratory (Griffiths)
> - Cardiovascular (Timmis/Robson)
>
> REAL-Health will create a series of provider-facing Discovery Information Utilities to support quality improvement programmes in east London including clinical decision support tools and pan-organisation GP practice performance dashboards. A Research Fellow and Programme Administrator have been appointed.

**Cardiovascular**

Preparatory work on atrial fibrillation work in the cardiovascular theme started in 2017 and has been accepted for publication.[2]

The relevant Discovery data extracts are currently being validated.

**Child Health**

National Child Measurement Programme data has been obtained and will be updated annually from Hackney, Newham and Tower Hamlets Local Authorities under signed Data Sharing Agreements (DSA). Initial analyses of data from these 3 CCGs comprising a cohort of over 50,000 children have been carried out and a paper submitted for publication.[3] Further analyses are underway to examine health outcomes for children with obesity. A Data Sharing Agreement with Waltham Forest is in preparation which will increase the geographical coverage of this work.

A Data Sharing Agreement has also been signed with the London Newborn Dried Blood Spot Programme, hosted at Great Ormond Street Hospital NHS Trust, to enable analysis of outcomes of newborn screening (including for sickle cell disorders) for North East London. This work has been approved by Public Health England and will assist in developing scalable national systems to assess quality and outcomes of this newborn screening programme.

**The SHARE project (Secure Health Analysis and Research in East London)**

SHARE, funded by Barts Charity, has established metadatasets for a number of research studies which have been approved by the Discovery Board including: East London Genes and Health (Van Heel, QMUL), dementia/Parkinson's disease (Noyce, Cuzick, Giavannoni, Barts Health/Wolfson Institute, QMUL), bowel surgery follow-up (Knowles, Barts Health/QMUL), diabetes outcomes and hypoglycaemia (Malawana, DeStavola (QMUL/UCL), childhood asthma, obesity and antibiotic use (Dezateux, Lyons, DeStavola; QMUL, Swansea, UCL), and antibiotic resistance and urinary tract infections (Hayward/Ayree UCL). These and other studies are awaiting Discovery data validation before obtaining Discovery access permissions and data extracts.

**East London Genes & Health (PI Van Heel)**

The project has currently recruited over 30,000 of the 100,000 expected participants and is progressing well.  Linkage of locally-registered patients is over 90%,: however 27% of current participants live outside the 3 CCG areas for whom CEG has currently accessible data, hence

---

[2] Robson J, Mathur R, Priebe M, Ahmed Z, Ayerbe L. Thromboembolic and haemorrhagic events in atrial fibrillation patients; a prospective cohort study. British Journal of General Practice.  (in press 2019)

[3] Firman N, Boomla K, Hudda M, Robson J,  Whincup, P, Dezateux C. Is child weight status correctly reported to parents? Analysis of National Child Measurement Programme data using ethnic-specific BMI adjustments.

accessing data from Discovery with a wider CCG inclusion and richer hospital data will be of substantial benefit to the study.

**Early Diagnosis Group (Peake, Robson et al)**

Lung cancer is the commonest cause of cancer death across north east London, with wide variation in treatment rates and survival. The underlying reasons for this remain unclear. This Cancer Research UK funded project is using linked primary and secondary care data to investigate this variation with a view to defining potential actionable moments allowing real time feedback to facilitate earlier referral with a view to reducing variance, increasing treatment rates and improving patient outcomes.

**HDRUK Implementation Projects**

These are designed to create nationally scaleable capability in electronic health record processing.

*Natural Language Processing*

Methods for natural language processing (NLP: the use of computational techniques to process and understand the written unstructured (not coded) elements of the electronic health record) are being explored through a national HDRUK initiative to develop UK-wide NLP capability using open source software (Gate; https://gate.ac.uk/) with proposals for a CEG-led use case on lung cancer pathology and imaging reports under consideration for funding. This complements work being developed through a Barts Health Life Science Initiative funded project based on diabetic foot ulcer disease in collaboration with Clinithink (Gutteridge, Boomla, Vijayagharavan, Majumdar, Claydon).

*Multimorbidity*

Methods for characterising and evaluating code sets to identify individuals with multimorbidity from electronic health records are being explored through a national HDRUK initiative (Robson and others) with funding for a CEG-led use case under consideration as a component. This will create a library of disease concepts that can be implemented in the Discovery Programme.

## *Publisher Usage*

Organisations that publish their data into Discovery have access to their own data, and (with appropriate governance agreements) to other publishers' data.

**Tower Hamlets Whole Systems Dataset Project**

> The Whole Systems Dataset Project, being carried out in conjunction with Tower Hamlets Local Authority, aims to establish an integrated health, social care and wider determinants of health dataset across the Borough for effective risk stratification and resource allocation and reduction of health inequalities using Discovery East London to identify socio-demographic determinants, health status and service usage. Linkage of a data extract for residents in local authority social care has been successfully achieved and is currently being analysed.

**Hurley Group**

> The Hurley Group is a consortium of around 16 GP practices and treatment centres in London using a common EMIS electronic record system. The Discovery database includes data from the Hurley Group giving Hurley Group practices access to identifiable patient data for their registered patients in order to:

> o  Create real-time dashboards with disease finding capability, visible at all the practices (e.g. number of possible diabetics); accessibility, demographics and disease burden, together with queries to help identify patients that are pre-deterioration or if a site needs support in an area that impacts health.

> o  Interrogate and interact with the data.

8.  **ONGOING SUSTAINABILITY**

**NEL STP-wide relationships**

Discussions are underway with local community, mental health and social care providers across the four London Boroughs.  Individual use cases are in development to test the scope and nature of the ongoing relationship between the London Boroughs and Discovery.

Work is underway with teams in West Essex and Barking, Havering & Redbridge with a view to extending the current partnership membership to make it coterminous with East London Health & Care Partnership geography: this includes discussions on opportunities for joint working and the infrastructure/system discussions required to create interoperability across the geographies.  Linked with this work, initial work with Care City in Barking is underway to identify potential links between the NHS test bed for innovation and Discovery East London as part of the work on Care City as a strategic priority for UCLPartners.

**'One London': Local Health and Care Record Exemplar (LHCRE)**

One London is one of five Local Health and Care Record Exemplars – or LHCREs – commissioned in 2018 by NHS England. They are designed to join up data from health and social care, enabling data to be accessed appropriately for defined purposes to deliver improved clinical and care outcomes and to provide measurable public benefit across whole system populations. The LHCREs are designed to be locally-led, embedding digital change as an enabler for service transformation on the ground and meaning that key clinical information will be instantly available to teams delivering care to the population. This ambition has been reinforced in the recently released NHS 10 year plan.

In London - and working as a partnership between the NHS, local authorities and the Greater London Authority - the One London model is designed to fully harness the power of data across this highly complex system to not only improve health outcomes, but also Londoners' experience of treatment in health and social care.

The Discovery Programme underpins a key component of the vision for One London and success in delivery and implementation across the NEL STP will provide leadership and confidence in scaling Discovery to other parts of London. CEG and Discovery contributed to the One London application which is one of the first five LHCREs to be awarded and the largest. This represents an important opportunity to develop a sustainable and larger vision for the Discovery Programme.

As part of the One London bid, demonstrator projects, selected through an open and transparent process, will be carried out. These are listed below and include a number which are led from, or contributed to, by north east London. This enables cross-STP working and ensures that utilities developed in some parts of London can be deployed in others, avoiding duplication and enabling harmonised approaches to common health and care functions.

- o Urgent & Emergency Care Pathway
- o Coordinate My Care: end of life care
- o Optimising health care for Londoners in > 500 care homes
- o Improving diagnostic care: radiology and pathology
- o Preventing stroke among 100,000 Londoners with atrial fibrillation
- o Integrating mental and physical health care
- o Whole population management of 600,000 Londoners with diabetes
- o Digital tools for self-management and patient engagement
- o Strategic planning of adult social care and wider determinants of health

As mentioned in section 6, One London includes a major workstream to establish on-going conversations with Londoners. The Discovery Programme will be contributing to, and aligning with, this programme to ensure that the messaging and communications reflect the views of our local communities while being consistent across London.

One London includes partnerships with the three London Academic Health Science Networks and the newly established London site of HDRUK, enabling the future impact of the Discovery

Programme to be maximised.  The Discovery Programme is contributing to the technical developments underpinning the One London vision with potential interaction with and leverage of sustainable infrastructure for cross-London working, including in digital innovations.

**DISCOVERY EAST LONDON**
**Frequently Asked Questions**

**January 2019**

---

### 1. What is the Discovery Programme?

A Programme to leverage a new data service set up to support direct care locally and support the development of a learning health information system that has benefits for direct quality of care and research to improve health at a population level. For robust analysis of population-level outcomes, a complete data set is often required.

It covers the 4 CCGs of TH, C&H, NH and WF and the Barts Health and Homerton NHS Trusts with a population coverage of ~1.5million (200 GP practices).

It's aim is to link up, subject to sharing agreements, data about patients to improve individual patient care. As an example: if a sharing agreement is set up between GP practice and school nursing service, then the GP practices would be able to view records of HPV Vaccine given by the School Nurses, by looking on the EMIS clinical system.

The secondary aim is to produce a depersonalised linked database similar to what CEG does with primary care data but extended now to include secondary care.

The project is focussed on combining the fractions of complete patient data held by different providers e.g. Secondary Care trusts and GPs, extracting all data including free text from every health and social care provider. As a first step, the project is focusing on data processing from local GP practices and hospitals (Barts, Homerton, Newham).

### 2. Who is funding the project?

The project is largely funded by the Endeavour Healthcare Charity but each of the four inner North East London CCGs (TH, CH, NH, WF) have contributed funds pooled between the organisations. The programme is hosted by the CEG.

### 3. What kind of data is extracted?

All data is extracted including free text. Free text in consultations could be useful for different specialities e.g. in diabetes it would be important for clinicians to know why patient stopped taking a drug and the reasons could have been written in free text. Free text also allows Discovery to explore the use of Natural Language Processing to help identify patient risk from detail included in the record's free text data.

### 4. How is this different from HIE?

HIE simply 'views' data from various providers in the sharing network. No aggregate information is held on servers, access is governed by a data sharing agreement and HIE data cannot be used for

analysis.  Discovery allows combined shared data to be held on its cloud-based server and this can be analysed to improve care.

**5.    Do we need to get patient consent?**

Patient consent is not required at this stage.  As Discovery is acting as data processor for the data controllers and dealing with patient data fed into the system.  The practice remains the data controller in the same way as the current systems (e.g. EMIS) arrangements.  Further details about consent will be provided in later stages of the project.

**6.    How is data shared and held?**

Data is extracted from clinical systems via direct application interfaces.  The data is held on Discovery servers providing secure access via NHS standards-based APIs.  Access is governed by NHS standard security model, including data sharing rules and access for users with specific roles.  The individual providers (e.g. GP practices) remain the primary Data Controllers and retain control over their data within Discovery at all times.

**7.    Who has access to the data?**

At this stage, access to the data is limited to the Discovery Programme Team and members of the CEG team who are involved in validating the accuracy of the data in Discovery.  Access is determined case-by-case by approved information sharing agreements between the data controllers and Discovery. Approvals are governed with involvement of data processors through the programme governance mechanisms described above.

**8.    How will access to the data work in practice?**

The Discovery Programme has a Board made up of IT / informatics leads from stakeholder organisations including CCGs, CEG, acute Trusts and mental health Trusts from across North East London. CEG approach the Discovery Board with any request to use the held datasets for projects initiated by any of the member CCGs, Trusts, related academic teams or broader health and social care providers across North East London.  Any request for access by a private company would also go to the Discovery Board for full review prior to approval.

Following approval by the Discovery Board, appropriate approved data may be transferred to approved secured server space with agreement for use by the approved applicant for a stated purpose.  This approach is in line with the access requirements set by external organisations such as NHS England or NHS Digital, for instance.

**9.    How does this fit in with the spine?**

All patients have their demographic details on the spine via the Patient Demographic Service. Discovery is not related to the spine in any way, currently. Were this to change in the future, practices would be informed.  No clinical data from Discovery passes to national systems and Discovery is not connected to the Summary Care Record.

If patients wish to opt out of the Summary Care Record (SCR), there are Read codes for GPs to enter within EMIS for this.  That blocks a data flow from EMIS to the SCR.  Signing up to Discovery does not affect this.

**10. How will patients opt out of having their records shared with other contributors to Discovery**

If patients do not want their records shared with other specific contributors through Discovery, that will be handled within the Consent Manager module of Discovery. Patients will have full control over who their records are shared with.

The process is:

- All patient data is replicated within the practice domain and under practice data controllership within Discovery.
- No data will flow into a specific organisation's domain unless the practice has a sharing agreement with that organisation.
- Patient dissent to this flow overrides all other considerations, so patient dissent will block this flow or indeed any other flow from the practice domain within Discovery.

**11. What control do patients have over their anonymised data?**

The Data Protection Act and the recommendations of the National Data Guardian, to be published as part of the Caldicott Review, regard anonymised aggregated data as being publishable. For example, data detail at the level of: *"There are 2,053 diabetics at your practice, 27% of whom smoke".* Discovery will reflect best practice in disclosure control; for example, as used by the Office of National Statistics to ensure patient confidentiality is protected.

**12. What about pseudonymised depersonalised data?**

Pseudonymised depersonalised data is data where the NHS number is replaced with pseudo ID, date of birth with age etc. If this is handled in a secure environment, this is not regarded in law as personal information. However, practices would obviously want to respect patient wishes on this use of their data, where possible. Dame Fiona Caldicott's recommendations on the use of this data has been published, consulted on and the government is about to respond. We await the final report and when the Discovery Board gets sight of the recommendations, it will decide how to best comply with this. The recommendations will outline exactly what rights patients have over their depersonalised data. For example, it is very likely that informing Cancer Registries etc. will be compulsory.

**13. How can a Data Sharing Agreement in EMIS be activated?**

*Click on EMIS web button > Configuration > Data Sharing Manage > My Agreement > Data Distribution > Developing a learning health system in the east of London 24037 > Activate My Agreement*