

London SDE/AIC Programme: Introduction and Proposed Use-Cases

Dr. Joe Zhang Prof. James Teo Dr. Jorge Cardoso
Jawad Chaudhry Sigal Hachlili

Version 0.5 (last updated 2024 Apr 7)

Introduction

The [London AI Centre](#) (AIC) has been commissioned as part of the London Secure Data Environment (SDE) programme for its latest phase: to extend AI technologies and analytics capabilities to stakeholders and data environments across London. This document summarises the latest state of planning for the programme, as an aid to internal and external stakeholders including Integrated Care Boards and the wider London NHS ecosystem.

What is the London SDE?

The London Secure Data Environment (SDE) is a pan-London NHS programme that is part of a national effort to enable secure and more powerful analytics for NHS, academic, and commercial users. Uniquely amongst regional peers, the London SDE does not focus on a single research platform. Rather, it places a focus on developing data infrastructure and capabilities that can support population health, care providers, and commissioners. This is in addition to building data environments that enable commercial research and development partnerships.

The SDE is led by **OneLondon**, as part of an overarching London Health Data Strategy, coalescing around three components (Figure [1](#)):

- (1) **London Data Service (LDS)**: hosted in North-East London, the LDS serves as a data engineering and service layer for pan-London primary care and secondary care data. It handles data extraction and linkage, and provisions data warehouses and secure analytics environments for both research and NHS users.

- (2) **DiscoverNOW Research/Analytics Environment:** run by Imperial College Healthcare Partners, DiscoverNOW supports governance and operation of secure research environments for academic, commercial, and NHS research and analytics.
- (3) **London AI Centre (AIC):** a national centre of excellence for applied data science and AI, the AIC provides frontier technology for data enrichment (CogStack), federated analytics (FLIP), and deployment of machine learning tools, as well as expertise in health data and advanced analytics.

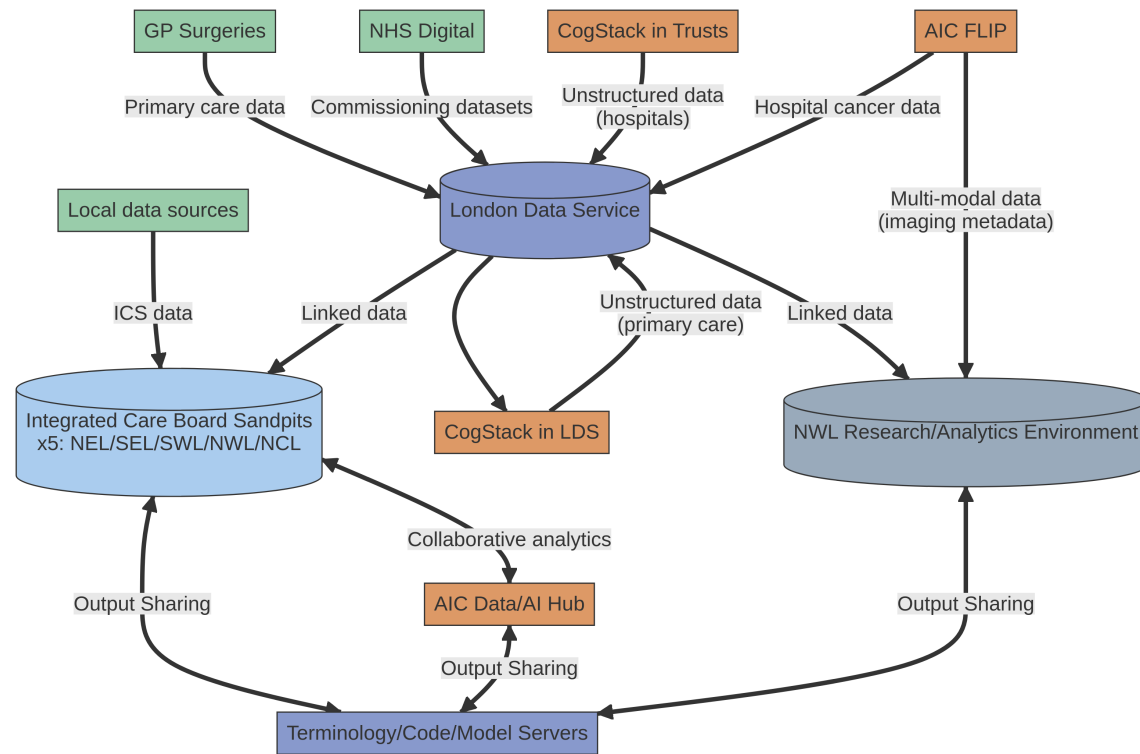


Figure 1: Summary of SDE components and data flows. Each London ICB is provisioned with its own data/analytics environment through the LDS. FLIP = Federated Learning and Interoperability Platform.

Source: [Article Notebook](#)

Technology and objectives

The contribution from the London AIC consists of technology deployment and supporting expertise, that enable a number of objectives (Figure 2) over the two year programme. This contribution includes the following:

- (1) **Federated Learning and Interoperability Platform (FLIP)**: Developed and tested over four years, FLIP consists of (a) secure data environments within NHS hospital Trusts for multi-modal imaging data, imaging metadata, and structured health record data in a common data model; and (b) a mechanism to query data and train AI models across these secure enclaves without the need to physically transfer data. FLIP is presently installed in four major London Trusts. Integrating FLIP into the SDE will enable hospital data (such as cancer data) to be surfaced into the LDS, and multi-modal capabilities to support research in precision healthcare.
- (2) **CogStack**: As an advanced natural language processing platform, CogStack can turn the large quantities of health information that are found in narrative text, into structured and analysable data. Currently actively used in Trusts to assist with clinical coding from notes and clinic letters, CogStack can surface secondary care and cancer pathway data, and previously unseen primary care data, into the SDE ecosystem.
- (3) **AIC Data/AI Hub**: The AIC hosts substantial health data and AI implementation expertise, that will provide practical support in data engineering, clinical informatics, data science, and machine learning (ML) development and deployment. Primary aims are to (a) help Integrated Care Boards (ICB) migrate data pipelines and analytics onto the LDS, (b) produce reproducible analytics pipelines for data science and predictive analytics capabilities, (c) work together to make ICBs self-sufficient in these capabilities.

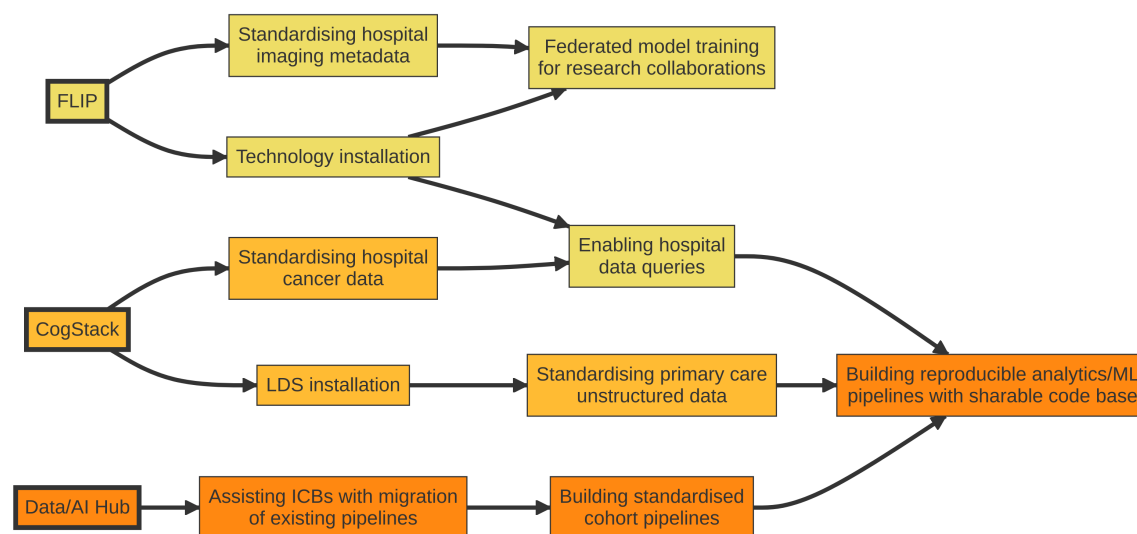


Figure 2: Summary of AIC work components and objectives. FLIP = Federated Learning and Interoperability Platform; ML = Machine Learning.

Source: [Article Notebook](#)

Of relevance to ICBs, resources are available to support migration of existing analytics into LDS Snowflake ‘Sandpits’, and constructing standard patient phenotype/cohort using definitions

hosted on a London terminology server. This will support building more complex reproducible analytics/machine learning pipelines, and delivery of ‘last mile’ insights to clinicians. As the LDS ICB environments share a common data model, any pipelines created in collaboration with a single ICB, can be adapted and used for any other ICB (or deployed across multiple environments to create pan-London insights). This will also facilitate shared terminologies, and validating/versioning/serving NHS-owned machine learning models across regions.

Proposed use-cases

The following use-cases are *examples* of analytics projects that can be supported within the SDE ecosystem, in collaboration between ICB/NHS analytics teams and the AIC/SDE team. Use-cases align to the London Health Data Strategy and long term condition priorities, as well as national programmes such as CORE20PLUS5, and are proposed here following early discussions with London ICBs. An overarching objective for any work is to build a code base that can be shared between ICBs and improved collaboratively.

Systematic measurement of group and individual health inequality

AIM: To systematically surface multiple dimensions of health inequality across sociodemographic/geospatial groups and individual patients, and to monitor this data continuously across key long-term conditions.

SUMMARY: Health inequality refers to measurable differences in health outcomes and determinants between individuals or groups (e.g. morbidity, co-morbidity, disease complications/death, healthcare access, disease screening, treatment delivery). Where individuals and groups experience health inequality, the principle of health *equity* emphasises the importance of reducing disparities by modifying outcome determinants that are unfairly distributed.

Health inequality is traditionally measured and visualised as a comparison of prevalence/incidence across different population groups. While helpful for broad insights, this offers limited understanding of complex individual circumstances. This type of measurement can be extended to individual patients, by using clinical domain knowledge to define ‘indicators’ of unequal disease, diagnosis, and treatment pathways. For example, in an individual with Diabetes Mellitus, indicators of inequality can include:

Diabetes surfacing at an early age;

Diagnosis in proximity to cardiovascular risk factor co-morbidities;

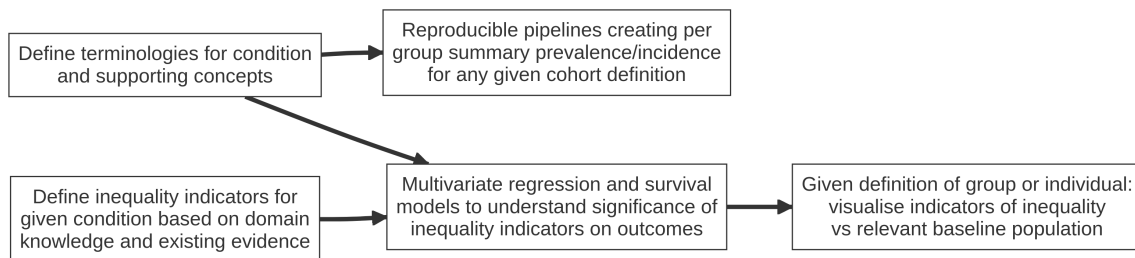
Diagnosis at a *late* age but with more severe disease, as measured by HbA1c or presence of end-organ complications;

Reduced health engagement/encounters/treatment compared to what is expected based on disease severity;

Shorter time to complications and mortality following diagnosis.

The precise contribution of factors to outcomes can be measured and understood in a multi-variate statistical model. Overall, the presence and magnitude of indicators can be used to visualise, monitor, and explain different types of inequality, including through comparison of groups and individuals to ‘what is expected’ in a background population. The outcome is an increase in actionability, with identification of modifiable determinants of inequality (= inequity) for small groups and individuals.

METHODS: The below shows an example workflow for Diabetes Mellitus, but can be applied to any long-term condition (not including cancer).



Source: [Article Notebook](#)

OUTPUTS: The primary output of this project would be a code-base that takes a cohort definition and a list of indicators as an input, and can be run to produce summary tables and statistics for groups and individual patients (where required). The code can be adapted by ICBs and used to support local dashboards. Code can be used for higher-level interval reporting and monitoring for the London region.

Cardiovascular disease prevention through decision intelligence

Actionable admission risk stratification

Joining up cancer pathways