**Criteria for lighthouse study selection**

*Niels Peek & Will Dixon*

*Version 1.3, 7th February 2023*

*GM SDE Programme, Work Stream 2*

**Background**: We aim to deliver a number of high priority, system-led research studies (“lighthouse studies”) to demonstrate the benefits of the GM Secure Data Environment (GM SDE). These lighthouse studies will be selected on agreed criteria, which are described in this document.

According to the *Secure Data Environments for Research and Development Expression of Interest Guide* published by NHS England in September 2022, SDEs for R&D should provide the infrastructure and analytical tooling to support the following types of research use cases:

1. AI/algorithm development (testing, training, and validation)
2. Clinical trial activities (feasibility, recruitment, efficacy through short- and long-term trial follow up)
3. Real world studies (safety, effectiveness, cost effectiveness)
4. Translational research (academic discovery and implementation of discovery into practice)
5. Epidemiological studies (large cohorts for population health research)
6. Health systems research (evaluation of systems or processes, including operational and applied research)

**Approach**: We expect that each of these six research use cases (henceforth called *study types*) will come with its own set of requirements – and therefore it is important to demonstrate capability for each of them. There will also be overlap in the requirements associated with different study types. For instance, each study type will require the availability of a secure research platform with state-of-the-art analytical tooling (e.g. R, Stata, Python).

We recognise that, regardless of study type, there exist several *additional requirements* that studies may have:

1. *Cohort linkage*
In order to conduct the study, a cohort dataset that is external to the GM SDE needs to be linked to data that are routinely available in the GM SDE
2. *Multi-modal data studies*
In order to conduct the study, longitudinal EHRs need to be linked other data modalities such images, signals, text, and omics data
3. *Projects with increased sensitivity*
Due to increased levels of sensitivity (e.g. involvement of commercial partners) it is imperative to use privacy-enhancing techniques
4. *Federated data analysis*
The study needs to be conducted across the footprint of GM, C&M, and L&SC utilising federated data analyis methods

In addition to covering the above listed study types A-F, we aim to demonstrate capability of the GM SDE for *data science training*, i.e., exposing data science students to the complexity of longitudinal EHRs for development of their analytical skills, and making them familiar with the technical infrastructure and data governance processes of a secure data environments.

**Selection criteria:**

Selection criteria a-b are aimed at covering the six study types and the four types of additional requirements. Selection criteria c-f focus at enhancing feasibilty of the selected LUCs and on minimising risks that they would not deliver.

1. Each selected lighthouse study is linked to one of the study types A-F listed above, or to data science training. Lighthouse studies may have one or more of the additional requirements i-iv.
2. The selected set of lighthouse studies covers all six study types A-F and all four additional requirements i-iv. The selected set of lighthouse studies therefore yields a comprehensive overview of required SDE functionalities.
3. Each lighthouse study is linked to an existing project with existing resources, a study team, and a skilled senior researcher that acts as principal investigator
4. The principal investigator and research team have the willingness and capacity to define, implement, and test the lighthouse study
5. For each lighthouse study there exist well-defined data specifications and a data analysis plan
6. For each lighthouse study there exists a legal basis for processing the data

**Next steps**: We will identify potential lighthouse studies for each study type, liaise with the study teams and principal investigators, and assess them against these criteria.